



**TUGAS AKHIR - SS141501**

## **PEMODELAN REGRESI POISSON INVERSE GAUSSIAN**

**(Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa  
Tengah Tahun 2015)**

**ANDRIANA YOSHINTA HERINDRAWATI  
NRP 1313 100 015**

**Dosen Pembimbing  
Dr. Drs. I Nyoman Latra, MS  
Dr. Purhadi, M.Sc**

**PROGRAM STUDI S1  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2017**



**TUGAS AKHIR - SS 141501**

**PEMODELAN REGRESI POISSON INVERSE  
GAUSSIAN**

**(Studi Kasus: Jumlah Kasus Baru HIV di  
Provinsi Jawa Tengah Tahun 2015)**

**ANDRIANA YOSHINTA HERINDRAWATI  
NRP 1313 100 015**

**Dosen Pembimbing  
Dr. Drs. I Nyoman Latra, MS  
Dr. Purhadi, M.Sc**

**PROGRAM STUDI S1  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2017**



**FINAL PROJECT - SS 141501**

**POISSON INVERSE GAUSSIAN REGRESSION  
MODELING**

**(Case Study: The Number of HIV New Cases  
in Central Java Province in 2015)**

**ANDRIANA YOSHINTA HERINDRAWATI  
NRP. 1313 100 015**

**Supervisor  
Dr. Drs. I Nyoman Latra, MS  
Dr. Purhadi, M.Sc**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2017**

**LEMBAR PENGESAHAN**

**PEMODELAN REGRESI POISSON INVERSE  
GAUSSSIAN  
(Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa  
Tengah Tahun 2015)**

**TUGAS AKHIR**

**Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Gelar  
Sarjana Sains  
pada  
Program Studi S-1 Jurusan Statistika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Institut Teknologi Sepuluh Nopember**

**Oleh:**

**ANDRIANA YOSHINTA HERINDRAWATI  
NRP. 1313100015**

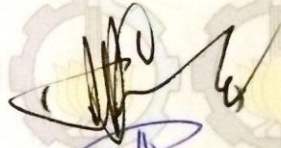

**Disetujui oleh Pembimbing Tugas Akhir**

**Dr. Drs. I Nyoman Latra, MS**

**NIP: 19511130 197901 1 001**

**Dr. Purnhadi, M.Sc**

**NIP: 19620204 198701 1 001**

()  
()

**Mengetahui**

**Ketua Jurusan Statistika FMIPA-ITS**



()

**Dr. Suhartono**

**NIP. 19710929 199512 1 001**

**SURABAYA, JANUARI 2017**

*(halaman ini sengaja dikosongkan)*

**PEMODELAN REGRESI POISSON INVERSE GAUSSIAN  
(Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa  
Tengah Tahun 2015)**

**Nama** : Andriana Yoshinta H.  
**NRP** : 1313 100 015  
**Jurusan** : Statistika FMIPA-ITS  
**Pembimbing** : Dr. Drs. I Nyoman Latra, MS  
: Dr. Purhadi, M.Sc

**Abstrak**

*Data count (data cacahan) adalah data yang menggambarkan jumlah kejadian pada kurun waktu tertentu dan hanya dapat bernilai positif, sehingga dalam pemodelannya dapat menggunakan regresi poisson. Terdapat asumsi yang harus dipenuhi jika menggunakan regresi poisson yaitu mean dan varians harus sama, sedangkan pada kasus data cacahan asumsi ini sering tidak terpenuhi. Hal ini terjadi karena adanya kasus overdispersi, yaitu varians lebih besar dari mean. Oleh karena itu dalam memodelkan data cacahan tersebut tidak cukup dengan regresi poisson sederhana. Regresi poisson inverse gaussian (PIG) merupakan salah satu bentuk regresi dari mixed poisson yang dirancang untuk data cacahan dengan kasus overdispersi dan telah digunakan pada beberapa penelitian yang menggunakan data cacahan. Penaksiran parameter dilakukan dengan metode MLE dan pengujian hipotesis dengan menggunakan metode MLTR. Jumlah kasus baru HIV merupakan salah satu data cacahan yang berpotensi terjadi overdispersi. Oleh karena itu, dalam memodelkan jumlah kasus baru HIV di Provinsi Jawa Tengah tahun 2015 dapat digunakan pemodelan dengan regresi PIG. Berdasarkan model tersebut, variabel prediktor yang memberikan pengaruh signifikan adalah persentase PUS yang menggunakan kondom, rasio fasilitas kesehatan, persentase daerah perkotaan, dan persentase penduduk usia 25-34 tahun.*

**Kata Kunci : Overdispersi, regresi Poisson Inverse Gaussian (PIG), HIV**

*(halaman ini sengaja dikosongkan)*

# **POISSON INVERSE GAUSSIAN REGRESSION MODELING**

**(Case Study: The Number Of HIV New Cases In Central Java  
Province In 2015)**

**Nama** : Andriana Yoshinta H.  
**NRP** : 1313 100 015  
**Jurusan** : Statistika FMIPA-ITS  
**Pembimbing** : Dr. Drs. I Nyoman Latra, MS  
: Dr. Purhadi, M.Sc

## **Abstract**

*Count data is data that describe a number of events at a certain time and can only be positive, so that it can be used Poisson Regression modeling. But there is assumption which should be fulfilled when using poisson regression, that is mean equals to variance, while in the case of data count, assumption of the data is often not fulfilled. This happens because overdispersion case, the variance is greater than the mean. Hence, modeling this kind of count data is not appropriate with the simple poisson regression. Poisson Inverse Gaussian (PIG) regression is one of regression formed by the mixed poisson distribution which is designed for overdispersion data. Parameter estimation in PIG regression is obtained with Maximum Likelihood Estimation (MLE) method and the statistical test is established with Maximum Likelihood Ratio Test (MLRT) method. The number of HIV new cases is one of the overdispersion count data. Therefore, PIG regression can be implemented for modeling the number of HIV new cases in Central Java Province in 2015. Based on the model, predictor variables which significant effect is the percentage of couples with childbearing age who use condom, ratio of health facilities, the percentage of urban areas, and the percentage of the population aged 25-34.*

**Key Word : Overdispersion, Poisson Inverse Gaussian Regression, HIV.**



*(halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, serta hidayah-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul

**“Pemodelan Regresi Poisson Inverse Gaussian  
(Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa  
Tengah Tahun 2015)”**

Tugas Akhir dapat terselesaikan dengan baik dan lancar tidak lepas dari banyaknya bantuan yang diberikan oleh berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada :

1. Dr. Drs. I Nyoman Latra, MS dan Dr. Purhadi, M.Sc selaku dosen pembimbing yang telah meluangkan waktu, mengarahkan, membimbing dengan sabar, dan memberikan dukungan yang sangat besar bagi penulis dalam menyelesaikan Tugas Akhir.
2. Dr. Suhartono selaku Ketua Jurusan Statistika ITS yang telah menyediakan fasilitas untuk menyelesaikan Tugas Akhir ini.
3. Dr. Sutikno, S.Si, M.Si selaku Ketua Prodi S1 Jurusan Statistika ITS.
4. Ir. Sri Pingit W., M.Si dan Dr. Wahyu Wibowo, S.Si, M. Si selaku dosen penguji yang telah memberikan saran-saran untuk kesempurnaan Tugas Akhir ini.
5. Ir. Sri Pingit W., M.Si selaku dosen wali yang telah memberikan nasehat, motivasi, serta bimbingan selama perkuliahan.
6. Ibu tersayang Indrawati atas doa, nasehat, kasih sayang, dan dukungan yang selalu diberikan kepada penulis sehingga menjadi termotivasi untuk menyelesaikan Tugas Akhir ini.
7. Fitrandi Firdaus yang selalu memberi dukungan, semangat, menghibur, serta memberikan banyak bantuan selama pengerjaan Tugas Akhir.
8. Teman-teman seperjuangan Σ24 yang mengajarkan kebersamaan dan rasa kekeluargaan selama masa perkuliahan ini.

9. Semua pihak yang telah memberikan dukungan yang tidak dapat penulis sebutkan satu persatu.

Pembuatan laporan Tugas Akhir ini masih jauh dari kesempurnaan, besar harapan dari penulis untuk menerima kritik dan saran yang berguna untuk perbaikan di masa mendatang. Semoga laporan Tugas Akhir ini bermanfaat.

Surabaya, Januari 2017

Penulis

## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b> .....	i
<b>TITLE PAGE</b> .....	iii
<b>LEMBAR PENGESAHAN</b> .....	v
<b>ABSTRAK</b> .....	vii
<b>ABSTRACT</b> .....	ix
<b>KATA PENGANTAR</b> .....	xi
<b>DAFTAR ISI</b> .....	xiii
<b>DAFTAR GAMBAR</b> .....	xv
<b>DAFTAR TABEL</b> .....	xvii
<b>DAFTAR LAMPIRAN</b> .....	xix
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan .....	5
1.4 Manfaat .....	5
1.5 Batasan Masalah .....	5
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Statistika Deskriptif .....	7
2.2 Distribusi Poisson .....	8
2.3 Distribusi <i>Inverse Gaussian</i> .....	9
2.4 Overdispersi .....	9
2.5 Distribusi PIG .....	10
2.6 Regresi PIG .....	16
2.7 Estimasi Parameter Regresi PIG .....	17
2.8 Pengujian Parameter .....	20
2.8.1 Uji Secara Serentak .....	20
2.8.2 Uji Secara Parsial .....	22
2.9 Uji Korelasi .....	23
2.10 Multikolinearitas .....	24

2.11 Akaike Information Criterion (AIC) .....	24
2.12 HIV .....	24
<b>BAB III METODOLOGI PENELITIAN</b>	
3.1 Sumber Data .....	27
3.2 Variabel Penelitian .....	27
3.3 Langkah Penelitian .....	29
<b>BAB IV ANALISIS DAN PEMBAHASAN</b>	
4.1 Karakteristik Data .....	31
4.1.1 Karakteristik Kasus Baru HIV di Jawa Tengah .....	31
4.1.2 Analisis Deskriptif Variabel Penelitian .....	33
4.2 Pemeriksaan Korelasi dan Multikolinearitas .....	35
4.3 Uji Overdispersi .....	37
4.4 Pemodelan Regresi Poisson Inverse Gaussian (PIG) .....	39
4.5 Pengujian Hipotesis .....	40
4.5.1 Pengujian Parameter Secara Serentak .....	40
4.5.2 Pengujian Parameter Secara Individu .....	41
4.6 Pemilihan Model Terbaik .....	42
<b>BAB V KESIMPULAN DAN SARAN</b>	
5.1 Kesimpulan .....	47
5.2 Saran .....	48
<b>DAFTAR PUSTAKA</b> .....	49
<b>LAMPIRAN</b> .....	53
<b>BIODATA PENULIS</b> .....	71

## DAFTAR GAMBAR

	Halaman
<b>Gambar 4.1</b> Perkembangan Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2011-2015 .....	32
<b>Gambar 4.2</b> Peta Jumlah Kasus Baru HIV Terbanyak di Provinsi Jawa Tengah Tahun 2015 .....	33

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

	Halaman
<b>Tabel 3.1</b> Struktur Data dalam Penelitian .....	29
<b>Tabel 4.1</b> Analisis Deskriptif Variabel Respon dan Variabel Prediktor .....	34
<b>Tabel 4.2</b> Koefisien Korelasi Variabel Respon dan Variabel Prediktor .....	36
<b>Tabel 4.3</b> Nilai VIF Variabel Prediktor .....	37
<b>Tabel 4.4</b> Pengujian Overdispersi .....	38
<b>Tabel 4.5</b> Estimasi Parameter Kemungkinan Model Regresi PIG .....	40
<b>Tabel 4.6</b> Pengujian Parameter Regresi PIG Secara Serentak	41
<b>Tabel 4.7</b> Pengujian Parameter Regresi PIG Secara Individu	42
<b>Tabel 4.8</b> Nilai AIC dari Model Regresi Poisson Inverse Gaussian .....	43
<b>Tabel 4.9</b> Penaksiran Parameter Model Regresi PIG pada Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2015 .....	43



*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1</b> Data Jumlah Kasus Baru HIV serta Faktor- Faktor yang Diduga Mempengaruhi di Provinsi Jawa Tengah Tahun 2015 .....	53
<b>Lampiran 2</b> Program R Pada Uji Overdispersi .....	55
<b>Lampiran 3</b> Hasil Program R Pada Uji Overdispersi.....	56
<b>Lampiran 4</b> Program R pada Regresi Poisson Inverse Gaussian (PIG) .....	57
<b>Lampiran 5</b> Hasil Program R pada Regresi Poisson Inverse Gaussian (PIG) .....	58
<b>Lampiran 6</b> Surat Pernyataan Pengambilan Data .....	69

*(Halaman ini sengaja dikosongkan)*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Data *count* (data cacahan) adalah data yang menggambarkan jumlah kejadian pada kurun waktu tertentu dan hanya dapat bernilai positif, karena suatu kejadian tidak mungkin bernilai negatif. Pada pemodelan dengan data cacahan akan melanggar asumsi regresi OLS (*Ordinari Least Square*) yaitu, error mengikuti distribusi normal (normalitas) dan memiliki sifat heterokedastisitas (varians konstan), sehingga pada data cacahan tidak dapat menggunakan regresi OLS (Widiari, 2016). Pemodelan data cacahan dalam perkembangannya memunculkan pemodelan data cacahan dengan *Generalized Linear Models* (GLMs). GLMs adalah generalisasi dari model regresi klasik atau regresi OLS dan terdapat metode analisis bagi data yang tidak memenuhi asumsi distribusi normal (De Jong dan Heller, 2008). Salah satu anggota keluarga GLMs dari distribusi poisson adalah regresi poisson.

Asumsi yang harus dipenuhi pada regresi poisson adalah nilai variasi variabel respon Y harus sama dengan nilai rata-ratanya (Myers, Montgomery, Vining dan Robinson, 2010). Pada analisis regresi Poisson dengan data diskrit biasanya terjadi pelanggaran asumsi tersebut, dimana nilai variansi lebih kecil daripada nilai rata-rata yang biasa disebut dengan *underdispersi* atau nilai variansi lebih besar dari nilai rata-rata yang disebut *overdispersi*. Tahun 1992, Cosul dan Famoye menyatakan bahwa terkadang ditemukan kasus *overdispersi* pada data *count*. Data cacahan biasanya memiliki nilai integer yang sangat besar dan banyak mengandung nilai nol, sehingga variansnya cukup besar (Zhu dan Joe, 2009). Apabila terjadi pelanggaran asumsi, maka kesimpulan yang dihasilkan tidak valid karena terjadi *underestimate* pada estimasi standar error. Cara mengatasi *overdispersi* adalah membentuk beberapa pemodelan yang merupakan kombinasi dari distribusi poisson dengan beberapa distribusi baik diskrit maupun

kontinu (*mixed poisson distribution*). Pada kombinasi distribusi poisson hanya beberapa distribusi yang sering digunakan dalam penelitian dikarenakan perhitungannya yang rumit. Salah satu *mixed poisson distribution* yang digunakan dalam penelitian adalah distribusi Poisson Inverse Gaussian (PIG) dengan random efek yang memiliki distribusi inverse gaussian. Distribusi ini pertama kali diperkenalkan oleh Holla pada tahun 1966 (Karlis dan Nikoloulopoulos, 2005).

Distribusi PIG merupakan bentuk dari distribusi *Siche* (SI) dengan dua parameter. SI disebut sebagai model yang lebih baik dari model binomial negatif, terutama untuk data yang *overdispersi*. Distribusi PIG memiliki fungsi likelihood yang *close form* dan perhitungannya lebih mudah, sehingga banyak penelitian dengan data cacahan menggunakan model ini (Stasinopoulos dan Rigby, 2007a). Tahun 2004 Shoukri, Asyali, dan Vandorp menggunakan model PIG pada data jumlah kasus penyakit mastitis pada sapi perah di Ontario Kanada dan menghasilkan kesimpulan bahwa regresi PIG merupakan model yang lebih baik dari model regresi binomial negatif. Pemodelan regresi PIG juga digunakan pada penelitian Zha, Lord, dan Zou (2014) pada kasus jumlah kecelakaan motor yang terjadi di Texas dan Washington, menunjukkan bahwa pemodelan regresi PIG lebih baik dari model regresi binomial negatif untuk memodelkan kasus jumlah kecelakaan motor di dua tempat tersebut. Tahun 2016 Widiari menggunakan model regresi PIG pada data jumlah kasus baru HIV di Jawa Timur Tahun 2013 dan menghasilkan kesimpulan bahwa variabel yang signifikan berpengaruh adalah persentase penduduk dengan pendidikan terakhir SLTA, persentase PUS yang menggunakan kondom, dan rasio fasilitas kesehatan.

Jumlah kasus baru HIV di suatu wilayah merupakan salah satu bentuk data cacahan sehingga dalam pemodelannya bisa menggunakan regresi Poisson. Data mengenai jumlah kasus HIV juga berpotensi terjadi *overdispersi* sehingga untuk mengatasi kasus tersebut diperlukan pemodelan regresi dengan *mixed poisson*

*distribution*. HIV merupakan penyakit menular yang disebabkan oleh infeksi *Human Immunodeficiency Virus*. Virus ini menyebar melalui cairan tubuh, dan menyerang sistem kekebalan tubuh, khususnya sel CD4 atau yang sering disebut sel T (CDC, 2016). Infeksi tersebut menyebabkan beberapa sel tubuh hancur sehingga penderita mengalami penurunan ketahanan tubuh dan tidak dapat melawan infeksi maupun penyakit lain.

Penyakit HIV adalah *new emergining diseases* serta menjadi pandemic di beberapa negara dalam beberapa waktu terakhir. Pandemic adalah epidemic yang terjadi dalam daerah yang luas dan mencakup populasi yang banyak, sedangkan yang dimaksud epidemic adalah mewabahnya penyakit pada daerah tertentu dalam jumlah yang melebihi batas normal. Pada tahun 2015 berdasarkan data *World Health Organization* (WHO) sebanyak 36,7 juta penduduk di dunia terinfeksi HIV dan 2,1 juta diantaranya merupakan kasus baru (WHO, 2016). HIV merupakan salah satu penyakit menular seksual kategori kronis yang awalnya termasuk ke dalam paradigma pembangunan *Millenium Development Goals* (MDGs) yang harus tercapai pada tahun 2015. Namun seiring berjalannya waktu hingga tahun 2015 target program MDGs untuk HIV belum tercapai, sehingga saat ini HIV termasuk dalam program *Sustainable Development Goals* untuk mengakhiri epidemi HIV di tahun 2030 (UNAIDS, 2016). Di Indonesia HIV termasuk ke dalam tiga besar penyakit dengan kasus terbanyak setelah TBC dan malaria. Jumlah HIV positif yang ada di masyarakat dapat diketahui melalui layanan *Voluntary, Counseling, and Testing* (VCT) (Dinkes Jateng, 2013).

Provinsi Jawa Tengah pada tahun 2015 menduduki peringkat ke 5 dengan jumlah kasus kumulatif HIV tertinggi (Damar, 2015). Jumlah kumulatif kasus HIV di provinsi Jawa Tengah hingga tahun 2015 mencapai 12.267 kasus. Angka tersebut lebih kecil dibandingkan dengan angka yang sebenarnya terjadi (fenomena gunung es), dikarenakan baru sebagian kecil saja yang dilaporkan

maupun yang terdeteksi. Jumlah kumulatif kasus HIV merupakan akumulasi dari beberapa tahun sebelumnya, karena kasus HIV tidak dapat disembuhkan sehingga kasusnya akan terus bertambah dan terakumulasi. Jumlah kasus baru HIV dapat menjadi patokan suatu daerah untuk melihat keberhasilan pemerintahannya dalam menghambat laju penyebaran kasus HIV. Jumlah kasus baru HIV di Jawa Tengah tahun 2015 mencapai 1.467 kasus, jumlah ini mengalami peningkatan jika dibandingkan dengan tahun 2014 sebanyak 1352 kasus (Dinkes Jateng, 2016). Program dan strategi yang dilakukan pemerintah dalam penurunan jumlah kasus HIV saat ini adalah dengan melakukan edukasi, penyuluhan melalui media masa, dan meningkatkan fasilitas kesehatan, namun berbagai faktor sosial ekonomi dapat menjadi hambatan dalam pemanfaatan fasilitas-fasilitas dan pelayanan kesehatan. Hal ini membuat upaya-upaya yang dilakukan pemerintah masih belum bisa menurunkan jumlah kasus HIV. Oleh karena itu perlu dilakukan penelitian untuk mengetahui faktor-faktor apa saja yang mempengaruhi jumlah kasus HIV di provinsi Jawa Tengah.

Penelitian mengenai kasus HIV diantaranya dilakukan oleh Ratnasari (2013) menggunakan regresi poisson bivariat dan menunjukkan bahwa faktor-faktor yang mempengaruhi jumlah kasus HIV dan AIDS di Provinsi Jawa Timur tahun 2011 adalah persentase penduduk pengguna kondom, persentase kelompok umur 25-29 tahun, persentase daerah berstatus desa, persentase penduduk tamat SMA dan persentase penduduk miskin di tiap kabupaten dan kota. Data jumlah kasus baru HIV di provinsi Jawa Tengah merupakan data *count* dan diduga mengalami *overdispersi* sehingga dalam pemodelannya digunakan regresi PIG.

## 1.2 Rumusan Masalah

Jumlah kasus baru HIV di provinsi Jawa Tengah di tahun 2015 mengalami peningkatan dari tahun sebelumnya. Oleh karena itu, perlu diketahui faktor-faktor apa saja yang signifikan

meningkatkan jumlah kasus HIV. Berdasarkan hal tersebut permasalahan yang akan dibahas dalam penelitian ini adalah faktor-faktor apa saja yang berpengaruh signifikan terhadap jumlah kasus baru HIV di Provinsi Jawa Tengah tahun 2015 melalui pemilihan model terbaik dari regresi Poisson Inverse Gaussian (PIG).

### **1.3 Tujuan**

Berdasarkan latar belakang dan rumusan masalah, maka tujuan penelitian ini adalah sebagai berikut :

1. Memodelkan faktor-faktor yang berhubungan dengan jumlah kasus baru HIV di Provinsi Jawa Tengah menggunakan pendekatan regresi Poisson Inverse Gaussian (PIG).
2. Menentukan faktor-faktor yang berpengaruh signifikan terhadap jumlah kasus baru HIV di Provinsi Jawa Tengah.

### **1.4 Manfaat**

Manfaat penelitian ini dapat digunakan oleh pemerintah Provinsi Jawa Tengah khususnya dinas-dinas dalam bidang kesehatan untuk menekan jumlah kasus baru HIV dan dapat dijadikan informasi sebagai pengambilan tindakan pencegahan untuk kedepannya.

### **1.5 Batasan Masalah**

Ruang lingkup pada penelitian ini dibatasi pada kasus baru jumlah HIV di Provinsi Jawa Tengah tahun 2015 yang merupakan Data Buku Saku Kesehatan Provinsi Jawa Tengah dengan model regresi Poisson Inverse Gaussian. Penaksiran parameter menggunakan metode *Maximum Likelihood Estimation* (MLE) yang diselesaikan dengan algoritma *Fisher Scoring*.



*(Halaman ini sengaja dikosongkan)*

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Statistika Deskriptif

Statistika deskriptif adalah metode statistika yang berfungsi untuk memberikan gambaran umum tentang penyajian sampel atau populasi. Analisis statistika deskriptif dapat diartikan sebagai metode yang berkaitan dengan mengumpulkan, mengolah, dan menyajikan data sehingga memberikan informasi yang berguna (Walpole, 1995). Data dapat disajikan dalam bentuk grafik atau tabel, untuk ukuran pemusatan data dan ukuran penyebaran data dideskripsikan secara numerik. Ukuran pemusatan data meliputi rata-rata, nilai tengah, dan modus sedangkan ukuran penyebaran data meliputi ragam dan standar deviasi (Walpole, 1995).

Rata-rata (*mean*) adalah nilai yang dapat digunakan untuk memberikan informasi atau gambaran umum dari sekumpulan data. Perhitungan rata-rata dengan menjumlahkan nilai pada data dibagi dengan banyaknya data tersebut. Rumus yang digunakan untuk menghitung rata-rata adalah sebagai berikut (Walpole, 1995).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ dimana } n = \text{banyak data} \quad (2.1)$$

Nilai maksimal adalah nilai yang paling tinggi atau besar dari sekumpulan data yang telah diurutkan. Nilai minimum adalah nilai yang paling rendah atau kecil dari sekumpulan data yang telah diurutkan. Ragam (*varians*) adalah ukuran yang digunakan untuk melihat seberapa besar penyimpangan data. Perhitungan ragam ditampilkan pada persamaan (2.2)(Walpole, 1995).

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (2.2)$$

## 2.2. Distribusi Poisson

Distribusi Poisson merupakan suatu distribusi untuk peristiwa yang probabilitas terjadinya kecil, dimana kejadian tergantung pada selang waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel prediktor saling independen. Selang waktu tersebut dapat berupa berapa saja panjangnya, misalnya semenit, sehari, seminggu, sebulan, bahkan setahun. Daerah tertentu yang dimaksudkan dapat berupa suatu garis, suatu luasan, suatu volume, atau mungkin sepotong bahan (Walpole, 1995).

Distribusi Poisson memiliki ciri-ciri sebagai berikut:

1. Banyaknya percobaan yang terjadi dalam suatu selang waktu atau suatu daerah tertentu, tidak tergantung pada banyaknya hasil percobaan yang terjadi pada selang waktu atau daerah lain yang terpisah.
2. Peluang terjadinya satu hasil percobaan selama suatu selang waktu yang singkat sekali atau dalam suatu daerah yang kecil. Sebanding dengan panjang selang waktu tersebut atau besarnya daerah tersebut dan tidak bergantung pada banyak hasil percobaan yang terjadi diluar selang waktu dan daerah tertentu.
3. Peluang bahwa lebih dari satu hasil percobaan akan terjadi dalam selang waktu yang singkat tersebut atau dalam daerah yang kecil tersebut dapat diabaikan.

Fungsi peluang untuk data berdistribusi Poisson bergantung pada parameter tunggal, yaitu rata-rata  $\mu$ . Fungsi peluangnya adalah sebagai berikut :

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \text{ untuk } y = 0, 1, 2, \dots \text{ dan } \mu > 0 \quad (2.3)$$

Dalam distribusi Poisson, rata-rata dan variansi adalah bernilai sama dan dapat dituliskan sebagai berikut:

$$E(Y) = Var(Y) = \mu \quad (2.4)$$

### 2.3. Distribusi *Inverse Gaussian*

Distribusi *Inverse Gaussian* merupakan distribusi kontinu dengan fungsi kepadatan mirip dengan distribusi gamma tetapi dengan kemencengan lebih besar dan keruncingan tajam. *Inverse Gaussian* mempunyai dua parameter dan fungsi kepadatan peluang yang dapat ditulis sebagai berikut:

$$f(y) = \left(2\pi y^3 \sigma\right)^{-0.5} e^{-(y-\mu)^2 / 2y\mu^2 \sigma^2}, \quad y > 0 \quad (2.5)$$

dengan rata-rata dan varians

$$E(Y) = \mu \text{ dan } Var(Y) = \sigma^2 \mu^3,$$

dan  $\sigma^2$  adalah parameter *disperse*. *Inverse Gaussian* digunakan pada keadaan dengan kemencengan yang ekstrim. Nama *inverse Gaussian* sendiri berasal dari fungsi kumulant yang memiliki hubungan invers dengan fungsi kumulant (logaritma natural dari fungsi MGF) distribusi normal/ distribusi Gaussian (De Jong dan Heller, 2008).

### 2.4. Overdispersi

Pada regresi Poisson terdapat asumsi yang harus terpenuhi yaitu *mean* dan *variance* variabel respon harus sama (*equidispersion*). Overdispersi pada regresi Poisson terjadi ketika varians dari variabel respons lebih besar dari rata-rata. Salah satu cara yang dapat dilakukan untuk mendeteksi ada atau tidaknya overdispersi pada variabel respon yang akan diteliti adalah dengan statistik uji Khi Kuadrat Pearson, yaitu sebagai berikut:

$$VT = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{\bar{y}} = (n-1) \frac{s^2}{\bar{y}} \quad (2.6)$$

Nilai ini sama dengan rasio varians terhadap rata-rata (*variance-to-mean ratio*), dimana seringkali disebut sebagai indeks disperse, dikalikan dengan  $n-1$ , dimana  $n$  adalah ukuran sampel. Apabila nilai indeks disperse kurang dari 1 maka dapat dikatakan terjadi underdispersi, sebaliknya overdispersi terjadi bila nilai indeks disperse lebih dari 1 (Karlis dan Xekalaki, 2000).

Uji statistik yang bisa digunakan untuk mendeteksi *overdispersi* pada suatu data adalah uji *overdispersi* yang dapat menggunakan *package* AER dari *software* R. Uji ini implementasi uji yang dikembangkan oleh Cameron dan Trivedi (1990). Uji ini mengikuti ide sederhana yaitu bahwa dalam model regresi poisson, varians sama dengan rata-rata yaitu  $\mu$  sehingga dalam uji ini hipotesis yang digunakan adalah

$$H_0 : \text{var}(Y) = \mu_i$$

$$H_1 : \text{var}(Y) = \mu_i + a \cdot g(.)$$

dimana  $g(.)$  merupakan suatu fungsi tertentu. Secara sederhana, bila nilai  $a = 0$  maka dapat dikatakan equidispersi, sebaliknya bila  $a > 0$  maka dapat dikatakan overdispersi. Nilai koefisien  $a$  dapat diperkirakan oleh regresi OLS dengan uji statistik yang digunakan adalah

$$T = \frac{1}{2} \sum_{i=1}^n \left\{ (Y_i - \hat{\mu}_i)^2 - Y_i \right\} \quad (2.7)$$

Dimana  $\hat{\mu}_i = \mu_i(\mathbf{x}_i; \hat{\boldsymbol{\beta}})$  dengan  $\hat{\boldsymbol{\beta}}$  merupakan estimasi  $\boldsymbol{\beta}$  dari MLE di bawah model poisson. Distribusi yang asimtotik terhadap  $T$  adalah distribusi normal standar di bawah hipotesis nol (Cameron dan Trivedi, 1990)

## 2.5. Distribusi PIG

Distribusi PIG merupakan salah satu distribusi *mixed* poisson. Bentuk dari distribusi *mixed* poisson tergantung dari distribusi pada random efek  $(v)$ .

Misalkan  $g(v)$  adalah fungsi kepadatan peluang dari  $v$  dan distribusi marginal untuk  $Y$  diperoleh dengan integral  $v_i$ .

$$P(Y = y | \mu) = \int f(y | \mu, v) g(v) dv \quad (2.8)$$

dengan

$$f(y|\mu, v) = e^{-v\mu} \frac{(\mu v)^y}{y!}$$

Untuk distribusi PIG,  $v$  pada persamaan (2.8) diasumsikan mengikuti distribusi *inverse* Gaussian dan memiliki fungsi kepadatan peluang yang dapat ditulis sebagai berikut:

$$g(v) = (2\pi v^3)^{-0.5} e^{-(v-1)^2/2\tau v}, v > 0 \quad (2.9)$$

dimana

$$\tau = \text{Var}(V), E(V) = 1$$

Distribusi PIG dilambangkan dengan PIG  $(\mu, \tau)$  diberikan oleh:

$$P(y|\mu, \sigma) = \left(\frac{2z}{\pi}\right)^{\frac{1}{2}} \frac{\mu^y e^{\frac{1}{\tau}} K_s(z)}{(z\tau)^y y!}, \quad (2.10)$$

dimana

$$s = y - \frac{1}{2} \text{ dan } z = \sqrt{\frac{1}{\tau^2} + \frac{2\mu}{\tau}}$$

$K_s(z) = K_{y-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{(2\mu\tau+1)}\right)$  adalah fungsi Bessel modifikasi jenis ketiga (Willmott, 1987).

Berikut adalah pembuktian dari proses *mixed* Poisson dari persamaan (2.8):

$$\begin{aligned} P(Y = y|\mu) &= \int_0^{\infty} f(y|\mu, v) g(v) dv \\ &= \int_0^{\infty} f(y|\mu, v) g(v) dv \\ &= \int_0^{\infty} e^{-v\mu} \frac{(\mu v)^y}{y!} g(v) dv \\ &= \int_0^{\infty} e^{-v\mu} \frac{(\mu v)^y}{y!} (2\pi v^3)^{-0.5} e^{-(v-1)^2/2\tau v} dv \end{aligned}$$

$$\begin{aligned}
&= \frac{\mu^y (2\pi\tau)^{-\frac{1}{2}}}{y!} \int_0^\infty e^{-v\mu} v^{y-\frac{3}{2}} e^{-\frac{(v^2-2v+1)}{2\tau v}} dv \\
&= \frac{\mu^y (2\pi\tau)^{-\frac{1}{2}}}{y!} \int_0^\infty e^{-v\mu} v^{y-\frac{3}{2}} e^{-\frac{v}{2\tau}} e^{\frac{1}{\tau}} e^{-\frac{1}{2\tau v}} dv \\
P(Y = y | \mu) &= \frac{\mu^y (2\pi\tau)^{-\frac{1}{2}} e^{\frac{1}{\tau}}}{y!} \int_0^\infty e^{-v\mu} v^{y-\frac{3}{2}} e^{-\frac{v}{2\tau}} e^{-\frac{1}{2\tau v}} dv \quad (2.11) \\
&= \frac{\mu^y (2\pi\tau)^{-\frac{1}{2}} e^{\frac{1}{\tau}}}{y!} \int_0^\infty v^{y-\frac{3}{2}} e^{-v\left(\mu+\frac{1}{2\tau}\right)-\left(\frac{1}{2\tau}\right)v^{-1}} dv \\
&= \frac{\mu^y (2\pi\tau)^{-\frac{1}{2}} e^{\frac{1}{\tau}}}{y!} \int_0^\infty e^{-\left(\frac{2\mu\tau+1}{2\tau}\right)v-\left(\frac{1}{2\tau}\right)v^{-1}} v^{y-\frac{3}{2}} dv
\end{aligned}$$

Berdasarkan tabel integral oleh Gardshteyn dan Ryzhik, 1980 seperti yang disebutkan dalam Shoukri, Asyali, Vandrop dan Kelton, 2004 bahwa

$$\int_0^\infty e^{-ax^h - bx^{-h}} x^{s-1} dx = \left(\frac{2}{h}\right) \left(\frac{b}{a}\right)^{\frac{s}{2h}} K_{\frac{s}{h}}(2\sqrt{ab})$$

Sehingga

$$\begin{aligned}
\int_0^\infty e^{-\left(\frac{2\mu\tau+1}{2\tau}\right)v-\left(\frac{1}{2\tau}\right)v^{-1}} v^{y-\frac{3}{2}} dv &= 2 \left(\frac{1}{2\mu\tau+1}\right)^{\frac{y-\frac{1}{2}}{2}} K_{y-\frac{1}{2}} \left(2\sqrt{\left(\frac{2\mu\tau+1}{4\tau^2}\right)}\right) \\
&= 2 \left(\frac{1}{2\mu\tau+1}\right)^{\frac{y-\frac{1}{2}}{2}} K_{y-\frac{1}{2}} \left(\frac{1}{\tau} \sqrt{2\mu\tau+1}\right)
\end{aligned}$$

kemudian persamaan tersebut disubstitusikan ke persamaan akhir dari persamaan (2.11) maka

$$\begin{aligned}
 P(Y = y | \mu) &= \frac{\mu^y (2\pi\mu)^{\frac{1}{2}} e^{\frac{1}{\tau}}}{y!} 2(2\mu\tau + 1)^{-\left(\frac{y-\frac{1}{2}}{2}\right)} K_{y-\frac{1}{2}}\left(\frac{1}{\tau} \sqrt{(2\mu\tau + 1)}\right) \\
 &= \frac{\mu^y e^{\frac{1}{\tau}}}{y!} \left(\frac{2}{\pi\mu}\right)^{\frac{1}{2}} (2\mu\tau + 1)^{-\left(\frac{y-\frac{1}{2}}{2}\right)} K_{y-\frac{1}{2}}\left(\frac{1}{\tau} \sqrt{(2\mu\tau + 1)}\right) \\
 &= \frac{\mu^y e^{\frac{1}{\tau}}}{y!} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{\left(\frac{1}{\tau} \sqrt{(2\mu\tau + 1)}\right)^{\frac{1}{2}}}{\left(\frac{\tau}{\tau} \sqrt{(2\mu\tau + 1)}\right)^y} K_{y-\frac{1}{2}}\left(\frac{1}{\tau} \sqrt{(2\mu\tau + 1)}\right) \\
 &= \frac{\mu^y e^{\frac{1}{\tau}}}{y!} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{\left(\sqrt{\frac{(2\mu\tau + 1)}{\tau^2}}\right)^{\frac{1}{2}}}{\tau^2 \left(\sqrt{\frac{(2\mu\tau + 1)}{\tau^2}}\right)^y} K_{y-\frac{1}{2}}\left(\frac{1}{\tau} \sqrt{(2\mu\tau + 1)}\right) \\
 &= \left(\frac{2z}{\pi}\right)^{\frac{1}{2}} \frac{\mu^y e^{\frac{1}{\tau}} K_y(z)}{(z\tau)^y y!}
 \end{aligned}$$

Distribusi PIG ditentukan oleh dua parameter yaitu rata-rata ( $\mu$ ) sebagai parameter lokasi dan parameter dispersi ( $\tau$ ) sebagai parameter bentuk. Probabilitas distribusi PIG dapat dihitung sebagai berikut:

$$P(Y = y | \mu) = \frac{\mu^y e^{\frac{1}{\tau}}}{y!} \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\mu\tau + 1)^{-\left(\frac{y-\frac{1}{2}}{2}\right)} K_{y-\frac{1}{2}}\left(\frac{1}{\tau} \sqrt{2\mu\tau + 1}\right)$$

Dalam Shoukri, dkk, 2004, fungsi Bessel memiliki beberapa properti penting yaitu diantaranya sebagai berikut:



$$K_{\frac{1}{2}}(a) = K_{-\frac{1}{2}}(a) = \left(\frac{\pi}{2a}\right)^{\frac{1}{2}} e^{-a} \quad \text{dan} \quad K_{\frac{3}{2}}(a) = \left(1 + \frac{1}{a}\right) K_{\frac{1}{2}}(a)$$

Dengan menggunakan properti-properti pada fungsi Bessel di atas maka fungsi probabilitas dari distribusi Poisson Inverse Gaussian (PIG) dapat dihitung sebagai berikut:

$$\begin{aligned} P(Y=0) &= e^{\frac{1}{\tau}} \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\mu\tau+1)^{\frac{1}{4}} K_{\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right) \\ &= e^{\frac{1}{\tau}} \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\mu\tau+1)^{\frac{1}{4}} \left(\frac{\pi}{\frac{2}{\tau}\sqrt{2\mu\tau+1}}\right)^{\frac{1}{2}} e^{-\frac{1}{\tau}\sqrt{2\mu\tau+1}} \\ &= e^{\frac{1}{\tau}} \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\mu\tau+1)^{\frac{1}{4}} \left(\frac{\pi\tau}{2}\right)^{\frac{1}{2}} (2\mu\tau+1)^{-\frac{1}{4}} e^{-\frac{1}{\tau}\sqrt{2\mu\tau+1}} \\ &= e^{\frac{1}{\tau}(1-\sqrt{2\mu\tau+1})} \\ P(Y=1) &= \mu e^{\frac{1}{\tau}} \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\mu\tau+1)^{\frac{1}{4}} K_{\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\mu\tau+1}\right) \\ &= \mu e^{\frac{1}{\tau}} \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (2\mu\tau+1)^{\frac{1}{4}} \left(\frac{\pi}{\frac{2}{\tau}\sqrt{2\mu\tau+1}}\right)^{\frac{1}{2}} e^{-\frac{1}{\tau}\sqrt{2\mu\tau+1}} \\ &= \mu(2\mu\tau+1)^{-\frac{1}{2}} e^{\frac{1}{\tau}(1-\sqrt{2\mu\tau+1})} \\ &= \mu(2\mu\tau+1)^{-\frac{1}{2}} P(Y=0) \end{aligned}$$

$$\begin{aligned}
P(Y=2) &= \frac{\mu^2 e^{\frac{1}{\tau}}}{2!} \left( \frac{2}{\pi\tau} \right)^{\frac{1}{2}} (2\mu\tau+1)^{-\frac{3}{4}} K_{\frac{3}{4}} \left( \frac{1}{\tau} \sqrt{2\mu\tau+1} \right) \\
&= \frac{\mu^2 e^{\frac{1}{\tau}}}{2!} \left( \frac{2}{\pi\tau} \right)^{\frac{1}{2}} (2\mu\tau+1)^{-\frac{3}{4}} \left( 1 + \frac{1}{\frac{1}{\tau} \sqrt{2\mu\tau+1}} \right) \left( \frac{\pi}{\frac{2}{\tau} \sqrt{2\mu\tau+1}} \right)^{\frac{1}{2}} e^{-\frac{1}{\tau} \sqrt{2\mu\tau+1}} \\
&= \frac{\mu^2 e^{\frac{1}{\tau}}}{2!} \left( \frac{2}{\pi\tau} \right)^{\frac{1}{2}} (2\mu\tau+1)^{-\frac{3}{4}} \left( 1 + \tau (2\mu\tau+1)^{-\frac{1}{2}} \right) \left( \frac{\pi\tau}{2} \right)^{\frac{1}{2}} (\sqrt{2\mu\tau+1})^{-\frac{1}{2}} e^{-\frac{1}{\tau} \sqrt{2\mu\tau+1}} \\
&= \frac{\mu^2 e^{\frac{1}{\tau}}}{2!} (2\mu\tau+1)^{-1} \left( 1 + \tau (2\mu\tau+1)^{-\frac{1}{2}} \right) e^{-\frac{1}{\tau} \sqrt{2\mu\tau+1}} \\
&= \frac{1}{2} P(Y=0) \frac{\mu^2}{(2\mu\tau+1)} \left( 1 + \tau (2\mu\tau+1)^{-\frac{1}{2}} \right) \\
&= \frac{2\mu\tau}{2\mu\tau+1} \left( \frac{1}{4} \right) \mu (2\mu\tau+1)^{-\frac{1}{2}} P(Y=0) + \frac{\mu^2}{(2\mu\tau+1)} \left( \frac{1}{2} \right) P(Y=0) \\
&= \frac{2\mu\tau}{2\mu\tau+1} \left( \frac{1}{4} \right) P(Y=1) + \frac{\mu^2}{(2\mu\tau+1)} \left( \frac{1}{2} \right) P(Y=0) \\
P(Y=3) &= \frac{\mu^3 e^{\frac{1}{\tau}}}{3!} \left( \frac{2}{\pi\tau} \right)^{\frac{1}{2}} (2\mu\tau+1)^{-\frac{5}{4}} K_{\frac{5}{4}} \left( \frac{1}{\tau} \sqrt{2\mu\tau+1} \right)
\end{aligned}$$

dengan cara yang sama seperti  $P(Y=2)$  maka diperoleh

$$P(Y=3) = \frac{2\mu\tau}{2\mu\tau+1} \left( \frac{3}{6} \right) P(Y=2) + \frac{\mu^2}{2\mu\tau+1} \left( \frac{1}{6} \right) P(Y=1)$$

sehingga dapat disimpulkan bahwa

$$P(Y=0) = e^{\frac{1}{\tau} (1 - \sqrt{2\mu\tau+1})}$$

$$P(Y=1) = \mu(2\mu\tau+1)^{-\frac{1}{2}} P(Y=0) \quad (2.12)$$

$$\begin{aligned}
P(Y=y) &= \frac{2\mu\tau}{2\mu\tau+1} \left( 1 - \frac{3}{2y} \right) P(y-1) + \frac{\mu^2}{2\mu\tau+1} \frac{1}{y(y-1)} P(y-2). \\
& \quad y = 2, 3, \dots
\end{aligned}$$

Hal ini menjadi

$$M(y) = (y+1) \frac{p(y+1)}{p(y)}$$

$$\text{Sehingga } \frac{K_{y+\frac{1}{2}}(z)}{K_{y-\frac{1}{2}}(z)} = (1+2\mu\tau)^{\frac{1}{2}} M(y)$$

Fungsi pembangkit peluang (*probability generating function*) untuk PIG diberikan oleh :

$$P(z) = \sum_{i=1}^n p(y) z^y = \exp \left( \tau^{-1} \left[ 1 - \{1 - 2\tau\mu(z-1)\}^{\frac{1}{2}} \right] \right)$$

Rata-rata untuk distribusi PIG adalah:

$$E(Y) = E\{E(Y | \mu\nu)\} = E(\mu\nu) = \mu \quad (2.13)$$

Varians untuk distribusi PIG adalah:

$$Var(Y) = Var\{E(Y | \mu\nu)\} + E\{Var(Y | \mu\nu)\} = \mu + \tau\mu^2 \quad (2.14)$$

Dean, Lawless dan Wilmot *dalam* Widiari, 2016 menyatakan Y merupakan sejumlah kejadian pada suatu unit observasi dalam suatu kurun waktu tertentu dan  $\mu$  adalah rata-rata kejadian tersebut. Sementara  $\tau$  merupakan parameter overdispersi dan sama dengan  $Var(v)$ , yang disebabkan karena terdapatnya heterogenitas atau keragaman yang berkaitan dengan unit observasi dengan karakter tertentu.

## 2.6. Regresi PIG

Model Regresi PIG seperti pada persamaan berikut (Purnamasari, 2016):

$$Y_i \sim PIG(\mu_i, \tau)$$

$$\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \text{ atau } \ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad (2.15)$$

dengan

$$\mathbf{x}_i^T = [1 \quad x_{1i} \quad x_{2i} \quad \cdots \quad x_{ki}]$$

$$\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_k]^T$$

dimana  $i = 1, 2, \dots, n$  menunjukkan nomor observasi.

Dengan fungsi kepadatan peluang sebagai berikut:

$$P(Y = y | \mathbf{x}_i; \boldsymbol{\beta}; \tau) = \left\{ \frac{e^{x_i^T \boldsymbol{\beta} y_i} e^{\frac{1}{\tau}}}{y_i!} \left( \frac{2}{\pi \tau} \right)^{\frac{1}{2}} \left( 2e^{x_i^T \boldsymbol{\beta}} \tau + 1 \right)^{-\frac{\left( y_i - \frac{1}{2} \right)}{2}} K_{s_i}(z_i) \right\}$$

## 2.7. Estimasi Parameter Regresi PIG

Persamaan (2.15) merupakan model regresi PIG dan parameter  $\boldsymbol{\beta}$  pada regresi PIG ditaksir dengan metode *maximum likelihood*. Langkahnya adalah dengan menentukan fungsi *likelihood* dari distribusi PIG sebagai berikut:

$$L(\boldsymbol{\beta}; \tau) = \prod_{i=1}^n P(Y = y_i | x_i; \boldsymbol{\beta}; \tau),$$

$$L(\boldsymbol{\beta}; \tau) = \prod_{i=1}^n \left\{ \frac{\mu_i^{y_i} e^{\frac{1}{\tau}}}{y_i!} \left( \frac{2}{\pi \tau} \right)^{\frac{1}{2}} (2\mu_i \tau + 1)^{-\frac{\left( y_i - \frac{1}{2} \right)}{2}} K_{s_i}(z_i) \right\} \quad (2.16)$$

Fungsi *likelihood* tersebut diubah kedalam bentuk logaritma natural ( $\ln$ ), sehingga persamaannya menjadi sebagai berikut:

$$l(\boldsymbol{\beta}; \tau) = \ln L(\boldsymbol{\beta}; \tau)$$

$$= \sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta} + \frac{n}{\tau} - \ln \left( \sum_{i=1}^n y_i! \right) + \frac{n}{2} \ln \left( \frac{2}{\pi} \right) - \frac{n}{2} \ln \tau - \sum_{i=1}^n \left( \frac{2y_i - 1}{4} \right) \ln (2\mathbf{x}_i^T \boldsymbol{\beta} + 1) + \sum_{i=1}^n \ln K_{s_i}(z_i) \quad (2.17)$$

Kemudian ditentukan turunan pertama dan turunan kedua terhadap parameter  $\boldsymbol{\beta}$  dan  $\tau$  dari persamaan (2.17).

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{\partial \left( \sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta} + \frac{n}{\tau} - \ln \left( \sum_{i=1}^n y_i! \right) + \frac{n}{2} \ln \left( \frac{2}{\pi} \right) - \frac{n}{2} \ln \tau - \sum_{i=1}^n \left( \frac{2y_i - 1}{4} \right) \ln (2\mathbf{x}_i^T \boldsymbol{\beta} + 1) + \sum_{i=1}^n \ln K_{s_i}(z_i) \right)}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^n y_i \mathbf{x}_i^T - \sum_{i=1}^n \left( \frac{2y_i - 1}{4} \right) \frac{1}{2\mathbf{x}_i^T \boldsymbol{\beta} + 1} 2\mathbf{x}_i^T = \mathbf{0} \end{aligned} \quad (2.18)$$

$$\begin{aligned} \frac{\partial l}{\partial \tau} &= \frac{\partial \left( \sum_{i=1}^n y_i \mathbf{x}_i^T \boldsymbol{\beta} + \frac{n}{\tau} - \ln \left( \sum_{i=1}^n y_i! \right) + \frac{n}{2} \ln \left( \frac{2}{\pi} \right) - \frac{n}{2} \ln \tau - \sum_{i=1}^n \left( \frac{2y_i - 1}{4} \right) \ln (2\mathbf{x}_i^T \boldsymbol{\beta} + 1) + \sum_{i=1}^n \ln K_{s_i}(z_i) \right)}{\partial \tau} = 0 \\ &= \frac{-n}{\tau^2} - \frac{n}{2\tau} = 0 \end{aligned} \quad (2.19)$$

Selanjutnya dari persamaan (2.18) ditentukan turunan  $\boldsymbol{\beta}$  yaitu

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} = \frac{\partial \left( \frac{\partial l}{\partial \boldsymbol{\beta}} \right)}{\partial \boldsymbol{\beta}} \quad \text{dan} \quad \tau \quad \text{yaitu} \quad \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \tau} = \frac{\partial \left( \frac{\partial l}{\partial \boldsymbol{\beta}} \right)}{\partial \tau} . \quad \text{Dari persamaan}$$

$$(2.19) \text{ juga ditentukan turunan terhadap } \tau \text{ yaitu } \frac{\partial^2 l}{\partial \tau^2} = \frac{\partial \left( \frac{\partial l}{\partial \tau} \right)}{\partial \tau} .$$

Apabila persamaan di atas persamaan implisit dan non linier dalam parameter  $\boldsymbol{\beta}$  dan  $\tau$  sehingga untuk mendapatkan taksiran dari parameter  $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta}^T & \tau \end{bmatrix}^T$ , fungsi dimaksimumkan dengan menggunakan *Fisher Scoring Algorithm*, dengan persamaan sebagai berikut (Widiari, 2016):

$$\hat{\boldsymbol{\theta}}_{(r+1)} = \hat{\boldsymbol{\theta}}_{(r)} + \mathbf{I}^{-1} \left( \hat{\boldsymbol{\theta}}_{(m)} \right) \mathbf{D} \left( \hat{\boldsymbol{\theta}}_{(m)} \right), \quad (2.20)$$

Dimana

$$\hat{\boldsymbol{\theta}} = \left( \hat{\boldsymbol{\beta}}^T, \hat{\tau} \right)^T$$

$$\mathbf{D}(\hat{\boldsymbol{\theta}}) = \left( \frac{\partial l}{\partial \hat{\tau}}, \frac{\partial l}{\partial \hat{\boldsymbol{\beta}}^T} \right)^T$$

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_{(m)}) = -E \left[ \mathbf{H}(\hat{\boldsymbol{\theta}}_{(m)}) \right]$$

$$\mathbf{H}(\hat{\boldsymbol{\theta}}_{(m)})_{(k+1)(k+1)} = \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{\tau}^2} & \frac{\partial^2 l}{\partial \hat{\tau} \partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}^T} \end{bmatrix}_{\theta=\theta_m}$$

$$\text{Sehingga } \mathbf{I}(\hat{\boldsymbol{\theta}}_{(m)}) = -E \begin{bmatrix} \frac{\partial^2 l}{\partial \hat{\tau}^2} & \frac{\partial^2 l}{\partial \hat{\tau} \partial \hat{\boldsymbol{\beta}}} \\ \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\tau}} & \frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}^T} \end{bmatrix}$$

Matriks hessian merupakan matriks yang berisi turunan kedua dari fungsi *likelihood* terhadap parameter  $\beta$  dan  $\tau$ . Adapun langkah-langkah *Fisher Scoring Algorithm* sebagai berikut:

1. Menentukan vektor awal parameter  $\hat{\boldsymbol{\theta}}_0$  dengan mengasumsikan data memenuhi model regresi linier berganda:

$$\mathbf{Y}^* = \beta_{0(0)} + x_{i1}\beta_{1(0)} + \dots + x_{ip}\beta_{p(0)} + \varepsilon_i \text{ dimana } i=1,2,\dots,n$$

dan dengan metode kuadrat terkecil diperoleh

$$\hat{\boldsymbol{\beta}}_{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}^*$$

2. Membentuk vektor gradien  $\mathbf{D}(\hat{\boldsymbol{\theta}}_0)$ .
3. Membentuk matriks hessian  $\mathbf{H}(\hat{\boldsymbol{\theta}}_{(0)})$ .
4. Membentuk matriks informasi *Fisher*  $\mathbf{I}(\hat{\boldsymbol{\theta}}_0)$ .
5. Memasukkan nilai  $\hat{\boldsymbol{\theta}}_{(0)}$  sehingga diperoleh vektor gradien  $\mathbf{D}(\hat{\boldsymbol{\theta}}_{(0)})$  dan matriks hessian  $\mathbf{H}(\hat{\boldsymbol{\theta}}_{(0)})$
6. Mulai dari  $m=0$  dilakukan iterasi pada  $\hat{\boldsymbol{\theta}}_{(r+1)} = \hat{\boldsymbol{\theta}}_{(r)} + \mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_{(m)}) \mathbf{D}(\hat{\boldsymbol{\theta}}_{(m)})$ , nilai  $\hat{\boldsymbol{\theta}}_{(m)}$  merupakan

sekumpulan penaksir parameter yang konvergen saat iterasi ke- $m$ .

7. Jika belum diperoleh penaksiran parameter yang konvergen saat iterasi ke- $m$ , maka dilanjutkan kembali ke langkah 5 hingga iterasi ke- $m+1$ . Iterasi akan berhenti apabila nilai dari  $\|\hat{\boldsymbol{\theta}}_{(m+1)} - \hat{\boldsymbol{\theta}}_{(m)}\| \leq \varepsilon$  dan  $\varepsilon > 0$  adalah bilangan yang sangat kecil.

*Fisher Scoring Algorithm* kemudian digunakan dalam dua algoritma yang digunakan dalam *Package Gamlss* yaitu algoritma Rigby Stasinopoulos (*RS algorithm*) dan algoritma Cole Green (*CG algorithm*). *RS algorithm* merupakan generalisasi dari algoritma yang digunakan Rigby dan Stasinopoulos (1996) dalam melakukan *fitting* terhadap *Mean and Dispersion Additive Models* (MADAM), dimana tidak menggunakan turunan silang dari fungsi *likelihood*. Sementara itu *CG algorithm* merupakan generalisasi dari algoritma yang digunakan oleh Cole dan Green (1992) dimana menggunakan turunan pertama, turunan kedua, dan turunan silang dari fungsi *likelihood* terhadap parameter distribusi (Stasinopoulos dan Rigby, 2007b). *RS algorithm* digunakan pada 10 iterasi pertama yang bertujuan untuk menstabilkan proses *fitting* dan kemudian merubah menjadi *CG algorithm* yang melanjutkan sampai 50 iterasi. *CG algorithm* membuat konvergen lebih cepat mendekati maksimum.

## 2.8. Pengujian Parameter

Pengujian parameter pada model PIGR dilakukan dengan menggunakan metode MLRT meliputi pengujian hipotesis secara serentak pada parameter  $\boldsymbol{\beta}$  serta pengujian parsial parameter  $\boldsymbol{\beta}$  dan  $\tau$ .

### 2.8.1 Uji Secara Serentak.

Parameter yang diuji pada pengujian ini mencakup seluruh parameter  $\boldsymbol{\beta}$  secara bersama-sama dengan hipotesis sebagai berikut:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$H_1$  : minimal ada satu  $\beta_l \neq 0$  dengan  $l = 1, 2, \dots, k$

Statistik uji yang digunakan adalah ukuran statistik *likelihood ratio* yang dibentuk dengan menentukan himpunan parameter di bawah populasi ( $\Omega$ ) yaitu  $\Omega = (\boldsymbol{\beta}, \tau)$  dan himpunan parameter di bawah  $H_0$  benar ( $\omega$ ) yaitu  $\omega = (\beta_{\omega 0}, \tau_{\omega})$ . Pada himpunan parameter di bawah populasi, dibentuk fungsi *likelihood* untuk model penuh (*saturated*) yang melibatkan seluruh variabel prediktor  $L(\Omega)$ . Sedangkan pada himpunan parameter di bawah  $H_0$  benar, dibentuk fungsi *likelihood* untuk model yang tidak melibatkan variabel prediktor  $L(\omega)$ . Kedua fungsi tersebut adalah sebagai berikut:

$$L(\hat{\Omega}) = \prod_{i=1}^n \left\{ \frac{\left( e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}} \right)^{y_i}}{y_i!} \left( \frac{2}{\pi \hat{\tau}} \right)^{\frac{1}{2}} \left( 2 \left( e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}} \right) \hat{\tau} + 1 \right)^{-\frac{\left( y_i - \frac{1}{2} \right)}{2}} K_{s_i}(z_i) \right\}$$

$$\ln(L(\hat{\Omega})) = \sum_{i=1}^n y_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \frac{n}{\hat{\tau}} - \ln \left( \sum_{i=1}^n y_i! \right) + \frac{n}{2} \ln \left( \frac{2}{\pi} \right) - \frac{n}{2} \ln \hat{\tau} - \sum_{i=1}^n \left( \frac{2y_i - 1}{4} \right) \ln \left( 2 \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \hat{\tau} + 1 \right) + \sum_{i=1}^n \ln K_{s_i}(z_i)$$
(2.21)

$$L(\hat{\omega}) = \prod_{i=1}^n \left\{ \frac{\left( e^{\hat{\beta}_0} \right)^{y_i}}{y_i!} \left( \frac{2}{\pi \hat{\tau}_{\omega}} \right)^{\frac{1}{2}} \left( 2 \left( e^{\hat{\beta}_0} \right) \hat{\tau}_{\omega} + 1 \right)^{-\frac{\left( y_i - \frac{1}{2} \right)}{2}} K_{s_i}(z_i) \right\}$$

$$\ln(L(\hat{\omega})) = \sum_{i=1}^n y_i \hat{\beta}_0 + \frac{n}{\hat{\tau}_{\omega}} - \ln \left( \sum_{i=1}^n y_i! \right) + \frac{n}{2} \ln \left( \frac{2}{\pi} \right) - \frac{n}{2} \ln \hat{\tau}_{\omega} - \sum_{i=1}^n \left( \frac{2y_i - 1}{4} \right) \ln \left( 2 \hat{\beta}_0 \hat{\tau}_{\omega} + 1 \right) + \sum_{i=1}^n \ln K_{s_i}(z_i)$$
(2.22)

Kedua fungsi *likelihood* tersebut dibandingkan dalam bentuk devians berikut:

$$G = -2 \ln \left( \frac{L(\hat{\omega})}{L(\hat{\Omega})} \right)$$

$$= 2 \left( \ln(L(\hat{\Omega})) - \ln(L(\hat{\omega})) \right)$$
(2.23)



Statistik G adalah pendekatan dari distribusi *chi square* dengan derajat bebas  $\nu$  sehingga kriteria pengujiannya adalah tolak  $H_0$  apabila  $G_{hit} > \chi^2_{(\alpha, \nu)}$  dimana  $\nu$  adalah derajat bebas yang diperoleh dari jumlah parameter di bawah populasi dikurangi jumlah parameter di bawah  $H_0$ .

### 2.8.2 Uji Secara Parsial

Apabila  $H_0$  ditolak pada pengujian serentak maka selanjutnya dilakukan pengujian hipotesis secara parsial.

Hipotesis untuk menguji signifikansi parameter  $\beta$ .

$$H_0 : \beta_l = 0$$

$$H_1 : \beta_l \neq 0 \text{ dengan } l = 1, 2, \dots, k$$

Statistik uji yang digunakan dalam pengujian signifikansi parameter  $\beta$  adalah sebagai berikut:

$$Z = \frac{\hat{\beta}_l}{SE(\hat{\beta}_l)} \quad (2.24)$$

Kriteria uji tolak  $H_0$  apabila  $|Z_{hit}|$  lebih besar dari nilai  $Z_{\alpha/2}$

dimana  $\alpha$  tingkat signifikansi yang digunakan  $SE(\hat{\beta}_l)$  merupakan elemen diagonal diperoleh dari elemen diagonal utama ke- $(m+2)$  dari matriks *varians* dan *covarians* yang diperoleh dari persamaan:

$$C\hat{\sigma}_v(\hat{\theta}) = -\left(\mathbf{H}^{-1}(\hat{\theta})\right)$$

Pengujian parameter  $\tau$

Hipotesis untuk menguji signifikansi parameter  $\tau$ .

$$H_0 : \tau = 0$$

$$H_1 : \tau \neq 0$$

Statistik uji yang digunakan:

$$Z = \frac{\hat{\tau}}{SE(\hat{\tau})} \quad (2.25)$$

$Z = \frac{\hat{\tau}}{SE(\hat{\tau})}$  Kriteria pengujianya adalah tolak  $H_0$  apabila  $|Z_{hit}|$

lebih besar dari nilai  $Z_{\alpha/2}$  dimana  $\alpha$  adalah tingkat signifikansi yang digunakan.

## 2.9. Uji Korelasi

Korelasi merupakan suatu indikator yang digunakan dalam hubungan linear antar dua variabel (Draper & Smith, 1992). Koefisien korelasi didefinisikan sebagai berikut.

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Nilai koefisien korelasi berkisar antara -1 sampai 1. Kedua variabel dikatakan memiliki hubungan yang erat bila nilai koefisien mendekati 1, baik itu positif maupun negatif. Namun bila nilai korelasi 0 menunjukkan bahwa kedua variabel tidak memiliki hubungan erat. Nilai korelasi yang positif menunjukkan adanya hubungan berbanding lurus diantara kedua variabel tersebut, sebaliknya hubungan yang berbanding terbalik terlihat dari nilai korelasi yang negative (Triola, 2010).

Pengujian hipotesisnya sebagai berikut:

$H_0$  : Tidak ada hubungan antara kedua variabel.

$H_1$  : Terdapat hubungan antar kedua variabel.

Statistik uji yang digunakan adalah

$$t = \frac{r_{x,y} \sqrt{n-2}}{\sqrt{1-r^2}} \quad (2.26)$$

Keputusan tolak  $H_0$  jika  $|t_{hit}| > t_{\left(\frac{\alpha}{2}, n-2\right)}$ .

## 2.10 Multikolinearitas

Pada pemodelan regresi, korelasi antara variabel-variabel prediktor disebut dengan multikolinearitas. Hal ini akan menjadi salah satu masalah dalam analisis regresi, karena bila beberapa prediktor berkorelasi maka akan sulit membedakan pengaruh dari masing-masing variabel. Salah satu cara untuk mendeteksi terjadinya multikolinearitas adalah dengan melihat nilai *Variance Inflation Factor* (VIF), yaitu nilai yang menggambarkan kenaikan varians dari dugaan parameter antar variabel prediktor. Persamaan VIF dapat ditulis sebagai berikut.

$$VIF = \frac{1}{1 - R_j^2} \quad (2.27)$$

Dimana  $R_j^2$  adalah koefisien determinasi antara  $x_i$  dengan variabel prediktor lainnya. Apabila nilai VIF lebih dari 10 maka dapat dikatakan terjadi kasus multikolinearitas (Setiawan dan Kusri, 2010).

## 2.11 Akaike Information Criterion (AIC)

*Akaike Information Criterion* (AIC) merupakan salah satu sarana dalam pemilihan model. AIC memperkirakan kualitas masing-masing model, relatif terhadap model lain. Misalkan  $L$  adalah nilai maksimum dari fungsi *likelihood* suatu model, dan  $k$  adalah jumlah parameter yang diestimasi dalam model tersebut, maka nilai AIC dari model tersebut adalah sebagai berikut:

$$AIC = 2k - 2\ln\left(L(\hat{\theta})\right) \quad (2.28)$$

Apabila diberikan beberapa model untuk sebuah set data, maka model yang lebih baik adalah model dengan AIC kecil (Akaike, 1978).

## 2.12 HIV

HIV merupakan penyakit menular yang disebabkan oleh infeksi *Human Immunodeficiency Virus*. Virus ini menyebar melalui cairan tubuh, dan menyerang sistem kekebalan tubuh, khususnya sel CD4 atau yang sering disebut sel T (CDC, 2016). Infeksi tersebut menyebabkan beberapa sel tubuh hancur sehingga penderita mengalami penurunan ketahanan tubuh dan tidak dapat melawan infeksi maupun penyakit lain. Sistem kekebalan dianggap defisien ketika sistem tersebut tidak dapat lagi menjalankan fungsinya memerangi infeksi dan penyakit- penyakit. Orang yang kekebalan tubuhnya defisien (Immunodeficient) menjadi lebih rentan terhadap berbagai ragam infeksi, yang sebagian besar jarang menjangkiti orang yang tidak mengalami defisiensi kekebalan. Penyakit-penyakit yang berkaitan dengan defisiensi kekebalan yang parah dikenal sebagai “infeksi oportunistik” karena infeksi-infeksi tersebut memanfaatkan sistem kekebalan tubuh yang melemah.

Jumlah HIV positif yang ada di masyarakat dapat diketahui melalui layanan *Voluntary, Counseling, and Testing (VCT)* (Dinkes Jateng, 2013). HIV menular melalui hubungan seksual yang tidak terlindungi (vaginal maupun anal) dan oral sex dengan orang yang terinfeksi, donor darah, dan berbagi jarum suntik yang sudah terkontaminasi. Hubungan seksual dengan penggunaan kondom dapat meminimalisir penyebaran virus HIV. Jumlah kasus baru HIV di Jawa Tengah dari tahun 2012 hingga tahun 2015 cenderung meningkat.

Kasus HIV di Jawa Tengah didominasi kelompok umur seksual aktif. Upaya pemerintah yang dilakukan untuk mengurangi penyakit HIV diantaranya adalah dengan penambahan fasilitas kesehatan dan tenaga kesehatan. Hal ini diharapkan mampu menekan angka kasus HIV. Seseorang mungkin terjangkit penyakit HIV salah satunya dikarenakan status ekonomi. Penderita dengan status ekonomi rendah cenderung mengalami pengobatan yang terbatas. HIV merupakan penyakit yang banyak tertanam dalam ketimpangan sosial ekonomi (Whiteside, 2002). Belenggu kemiskinan membuat seseorang cenderung memilih pekerjaan

yang beresiko terjangkit HIV contohnya pekerja seks komersil. Penyakit HIV juga dipengaruhi oleh tingkat pendidikan, hasil survei SDKI tahun 2012 menyatakan bahwa tingkat pengetahuan akan HIV dari setiap orang berbeda menurut pendidikan yang ditamatkan. Seseorang dengan pendidikan SLTA ke atas cenderung lebih banyak tahu mengenai HIV/AIDS dan cara pencegahannya.

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Sumber Data**

Penelitian ini menggunakan data sekunder tentang jumlah kasus baru HIV dan faktor-faktor yang mempengaruhi jumlah kasus baru HIV. Data diambil dari Profil Kesehatan Provinsi Jawa Tengah dan Buku Saku Kesehatan 2015 yang dikeluarkan Dinas Kesehatan Provinsi Jawa Tengah dan Publikasi hasil survey Sosial Ekonomi Nasional (Susenas) yang dikeluarkan oleh Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Data tersebut merupakan data pada tahun 2015 dengan unit penelitian yang diambil pada tingkat Kabupaten/ Kota di Provinsi Jawa Tengah dengan 29 Kabupaten dan 6 Kota.

#### **3.2 Variabel Penelitian**

Variabel penelitian yang digunakan dalam penelitian tugas akhir ini terdiri atas variabel respon (Y) dan variabel prediktor (X). Variabel penelitian yang digunakan dalam tugas akhir ini antara lain.

1. Jumlah kasus baru HIV di setiap kabupaten atau kota di Provinsi Jawa Tengah tahun 2015 (Y).
2. Persentase penduduk miskin di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015 ( $X_1$ ).  
Perbandingan antara jumlah penduduk dengan pengeluaran perkapita di bawah garis kemiskinan (GK) terhadap jumlah total penduduk dikalikan 100 di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015.
3. Persentase penduduk dengan pendidikan tertinggi SMA di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015 ( $X_2$ ).  
Perbandingan antara jumlah penduduk usia 10 tahun ke atas dengan pendidikan tertinggi SLTA terhadap jumlah total

penduduk usia 10 tahun dikalikan 100 di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015.

4. Persentase pasangan usia subur (PUS) yang sedang menggunakan alat KB kondom di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015 ( $X_3$ ).  
Perbandingan antara jumlah pasangan usia subur (PUS) yang sedang menggunakan alat KB Kondom terhadap jumlah total pasangan usia subur dikalikan 100 di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015.
5. Rasio jumlah tenaga kesehatan per 100.000 penduduk di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015 ( $X_4$ ).  
Perbandingan antara jumlah tenaga kesehatan terhadap jumlah total penduduk dikalikan 100.000 di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015.
6. Rasio jumlah fasilitas kesehatan per 100.000 penduduk di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015 ( $X_5$ ).  
Perbandingan antara jumlah fasilitas kesehatan terhadap jumlah penduduk dikalikan 100.000 di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015.
7. Persentase daerah perkotaan di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015 ( $X_6$ ).  
Perbandingan antara jumlah desa/kelurahan yang berstatus daerah perkotaan terhadap jumlah total desa/elurahan dikalikan 100 di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015.
8. Persentase penduduk usia 25-34 tahun di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015 ( $X_7$ ).  
Perbandingan antara jumlah penduduk usia 25-34 tahun terhadap jumlah penduduk total dikalikan 100 di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2015.

Struktur data untuk variabel penelitian ini ditunjukkan oleh Tabel 3.1.

**Tabel 3.1** Struktur Data dalam Penelitian

Wilayah	Y	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
1	y <sub>1</sub>	$x_{1,1}$	$x_{2,1}$	$x_{3,1}$	$x_{4,1}$	$x_{5,1}$	$x_{6,1}$	$x_{7,1}$
2	y <sub>2</sub>	$x_{1,2}$	$x_{2,2}$	$x_{3,2}$	$x_{4,2}$	$x_{5,2}$	$x_{6,2}$	$x_{7,2}$
3	y <sub>3</sub>	$x_{1,3}$	$x_{2,3}$	$x_{3,3}$	$x_{4,3}$	$x_{5,3}$	$x_{6,3}$	$x_{7,3}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	y <sub>35</sub>	$x_{1,35}$	$x_{2,35}$	$x_{3,35}$	$x_{4,35}$	$x_{5,35}$	$x_{6,35}$	$x_{7,35}$

### 3.3 Langkah Penelitian

Langkah-langkah untuk menentukan faktor-faktor yang memengaruhi jumlah kasus baru HIV dengan model regresi Poisson Inverse Gaussian adalah sebagai berikut:

1. Membuat analisis statistika deskriptif untuk jumlah kasus baru HIV beserta faktor-faktor yang berpengaruh dengan menggunakan *mean*, median, dan modus.
2. Menguji korelasi antara variabel respon dengan variabel prediktor dengan uji T.
3. Melakukan pemeriksaan kasus multikolinieritas dengan menggunakan kriteria VIF, untuk mengetahui apakah antar variabel yang diduga berpengaruh terdapat hubungan yang erat.
4. Melakukan uji overdispersi.
5. Melakukan pemodelan dengan menggunakan regresi Poisson Inverse Gaussian yang meliputi.
  - a. Mendapatkan penduga parameter model regresi Poisson Inverse Gaussian dengan menggunakan *Maximum Likelihood Estimation* (MLE).
  - b. Melakukan pengujian hipotesis untuk regresi Poisson Inverse Gaussian menggunakan *Maximum Likelihood Ratio Test* (MLRT).



6. Membandingkan nilai AIC untuk mencari model terbaik.
7. Melakukan interpretasi model regresi Poisson Inverse Gaussian yang didapatkan.
8. Membuat kesimpulan dari hasil analisis tersebut.

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

Pada bab ini akan dijelaskan beberapa hal yang dilakukan untuk mencapai tujuan menjawab permasalahan dalam penelitian ini. Hal-hal yang akan dibahas diantaranya meliputi karakteristik data dengan analisis statistika deskriptif untuk data kasus baru HIV di Jawa Tengah tahun 2015, deteksi overdispersi, deteksi multikolinearitas, analisis regresi Poisson Inverse Gaussian untuk mencari faktor-faktor yang berpengaruh terhadap jumlah kasus baru HIV di Provinsi Jawa Tengah. Regresi Poisson Inverse Gaussian (PIG) merupakan regresi yang dapat diaplikasikan pada data *count* yang mengalami kasus overdispersi. Pada penelitian ini pemodelan regresi Poisson Inverse Gaussian diaplikasikan pada data jumlah kasus baru HIV di Provinsi Jawa Tengah pada tahun 2015 karena data tersebut mengalami kasus overdispersi.

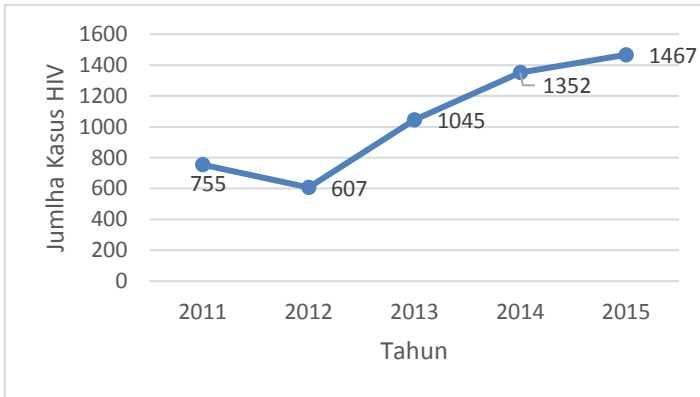
#### **4.1 Karakteristik Data**

Sebelum melakukan pengujian pada data, dilakukan terlebih dahulu analisis deskriptif dari variabel respon dan variabel prediktor untuk diketahui karakteristik dari masing-masing variabel yang akan digunakan pada penelitian tugas akhir ini.

##### **4.1.1 Karakteristik Kasus Baru HIV di Jawa Tengah**

Jawa Tengah adalah salah satu provinsi di wilayah pulau Jawa dengan luas 32548 km<sup>2</sup>. Provinsi Jawa Tengah memiliki 35 kabupaten/kota dengan rincian 29 kabupaten dan 6 kota. Pada setiap kabupaten/kota memiliki jumlah kasus baru HIV yang berbeda. Pada akhir bulan Desember tahun 2015 tercatat jumlah total kasus baru HIV yang ditemukan di setiap kabupaten/kota melalui Volunteer Conseling and Test (VCT) adalah sebanyak 1467. Angka temuan tersebut sesungguhnya jauh lebih kecil

dibandingkan angka sebenarnya (fenomena gunung es). Provinsi Jawa Tengah merupakan provinsi kelima setelah Provinsi Jawa Barat dengan jumlah kasus HIV terbanyak ditemukan. Perkembangan jumlah kasus baru HIV di Provinsi Jawa Tengah dari tahun 2011 hingga tahun 2015 dapat dilihat pada Gambar 4.1.



**Gambar 4.1** Perkembangan Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2011-2015

Berdasarkan Gambar 4.1 di atas dapat dilihat perkembangan kasus baru HIV dari tahun 2009 hingga tahun 2015. Pada tahun 2012 jumlah kasus HIV di Provinsi Jawa Tengah menurun sebanyak 148 kasus, namun pada tahun 2013 jumlah kasus baru HIV meningkat menjadi 1045 kasus. Di tahun 2013 hingga tahun 2015 jumlah kasus HIV di Provinsi Jawa Tengah mengalami peningkatan setiap tahunnya. Banyaknya kasus HIV yang ditemukan tidak terlepas dari peran klinik VCT. Pada Gambar 4.2 terlihat beberapa kabupaten/kota di Provinsi Jawa Tengah dengan jumlah kasus baru HIV terbanyak ditemukan diantaranya adalah Kabupaten Grobogan, Kabupaten Pemalang, Kabupaten Jepara, Kabupaten Batang, dan Kota Semarang. Kota Semarang merupakan kota dengan jumlah kasus baru HIV ditemukan

Variabel bebas yang digunakan dalam penelitian ini antara lain adalah persentase penduduk miskin di tiap kabupaten/kota Provinsi Jawa Tengah ( $X_1$ ), persentase penduduk dengan pendidikan tertinggi SLTA di tiap kabupaten/kota Provinsi Jawa Tengah ( $X_2$ ), persentase pasangan usia subur yang sedang menggunakan alat KB kondom di tiap kabupaten/kota Provinsi Jawa Tengah ( $X_3$ ), rasio jumlah tenaga kesehatan per 100.000 penduduk di tiap kabupaten/kota Provinsi Jawa Tengah ( $X_4$ ), rasio jumlah fasilitas kesehatan per 100.000 penduduk di tiap kabupaten/kota Provinsi Jawa Tengah ( $X_5$ ), persentase daerah perkotaan di tiap kabupaten/kota Provinsi Jawa Tengah ( $X_6$ ), dan persentase penduduk usia 25-34 tahun di tiap kabupaten/kota Provinsi Jawa Tengah ( $X_7$ ). Nilai dari masing-masing variabel dapat dilihat pada Lampiran 1.

Pada Tabel 4.1 dapat dilihat bahwa rata-rata jumlah kasus HIV di Provinsi Jawa Tengah tahun 2015 adalah 41,91 kasus setara dengan 42 kasus. Jumlah kasus tertinggi sebanyak 116 kasus di Kota Semarang dan terendah sebanyak 3 kasus di Kota Tegal. Rata-rata persentase penduduk miskin di Provinsi Jawa Tengah sebesar 13,031 persen dengan Kota Semarang memiliki persentase terendah dan Kabupaten Magelang memiliki persentase tertinggi. Jika dilihat dari segi pendidikan terakhir yang ditamatkan, rata-rata penduduk Provinsi Jawa Tengah yang memiliki pendidikan terakhir SLTA sebesar 12,340 persen dengan persentase terendah Kabupaten Wonosobo dan Kota Surakarta memiliki persentase tertinggi.

**Tabel 4. 1** Analisis Deskriptif Variabel Respon dan Variabel Prediktor

Variabel	Mean	Stdev	Variance	Minimum	Maximum
Y	41,91	26,45	699,85	3,00	116,00
X <sub>1</sub>	13,031	4,315	18,618	4,970	21,450
X <sub>2</sub>	12,340	5,309	28,182	6,680	25,230
X <sub>3</sub>	2,224	1,635	2,673	0,350	7,410
X <sub>4</sub>	33,89	42,76	1828,63	5,13	222,72
X <sub>5</sub>	3,863	1,591	2,533	2,200	9,940
X <sub>6</sub>	21,74	36,25	1314,27	1,23	100,00
X <sub>7</sub>	7,468	0,667	0,446	5,820	9,090

Rata-rata persentase pasangan usia subur (PUS) yang sedang menggunakan alat KB Kondom sebesar 2,224 persen dengan Kabupaten Kendal memiliki persentase terendah dan Kota Surakarta memiliki persentase tertinggi. Rasio jumlah tenaga kesehatan per 100.000 penduduk memiliki rata-rata sebesar 33,89 dengan Kabupaten Cilacap memiliki rasio terendah dan Kota Magelang memiliki rasio tertinggi. Rata-rata rasio jumlah fasilitas

kesehatan per 100.000 penduduk sebesar 3,863 dengan rasio terendah pada Kabupaten Sukoharjo dan rasio tertinggi pada Kota Magelang. Daerah di Provinsi Jawa Tengah yang berstatus daerah perkotaan memiliki rata-rata sebesar 21,74 persen, dengan Kabupaten Pati yang memiliki daerah berstatus perkotaan terendah. Rata-rata persentase penduduk usia 25-34 tahun sebesar 7,468 persen dengan Kabupaten Wonogiri memiliki persentase terendah dan Kota Semarang memiliki persentase tertinggi.

#### **4.2 Pemeriksaan Korelasi dan Multikolinearitas**

Pemeriksaan hubungan antara variabel respon dengan variabel prediktor dilakukan terlebih dahulu sebelum melakukan pemodelan dengan regresi Poisson Inverse Gaussian (PIG). Hal ini dapat dilakukan dengan analisis korelasi. Analisis korelasi adalah suatu metode analisis statistik yang digunakan untuk mengukur kekuatan hubungan antara dua variabel melalui suatu bilangan yang disebut koefisien korelasi. Selain itu, hal lain yang perlu dilakukan adalah pengecekan hubungan antara variabel prediktor yang bertujuan untuk mengecek apakah terjadi multikolinearitas atau tidak. Pengujian multikolinearitas dilakukan agar nantinya penduga parameter dari variabel prediktor tidak bias.

Pengujian hipotesis sebagai berikut:

$H_0$  : Tidak ada hubungan antar kedua variabel

$H_1$  : Terdapat hubungan antar kedua variabel

Koefisien korelasi antar variabel respon dengan variabel prediktor serta antara variabel prediktor disajikan pada Tabel 4.2. Berdasarkan Tabel 4.2 terlihat bahwa dengan  $\alpha = 0,1$  beberapa prediktor yang memiliki hubungan dengan variabel respon antara lain rasio jumlah tenaga kesehatan per 100.000 penduduk dan rasio jumlah fasilitas kesehatan per 100.000 penduduk. Sementara itu, variabel persentase penduduk miskin, persentase penduduk dengan pendidikan terakhir SLTA, persentase pasangan usia subur (PUS)

yang menggunakan kondom, dan persentase penduduk usia 25-34 tahun tidak memiliki hubungan dengan variabel respon. Akan tetapi, variabel-variabel tersebut tetap dimasukkan dalam analisis karena secara teori variabel-variabel tersebut berpengaruh secara signifikan terhadap variabel respon.

**Tabel 4. 2** Koefisien Korelasi Variabel Respon dan Variabel Prediktor

Korelasi	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
X <sub>1</sub>	-0,001 0,997						
X <sub>2</sub>	-0,088 0,614	-0,692 0,000					
X <sub>3</sub>	-0,037 0,834	-0,420 0,012	0,766 0,000				
X <sub>4</sub>	-0,315 0,066	-0,467 0,004	0,772 0,000	0,782 0,000			
X <sub>5</sub>	-0,460 0,005	-0,336 0,048	0,596 0,000	0,718 0,000	0,850 0,000		
X <sub>6</sub>	-0,229 0,186	-0,565 0,000	0,799 0,000	0,826 0,000	0,814 0,000	0,767 0,000	
X <sub>7</sub>	0,184 0,290	-0,523 0,001	0,613 0,000	0,395 0,019	0,325 0,057	0,123 0,480	0,574 0,000

Pemeriksaan adanya multikolinearitas antar variabel prediktor dilakukan karena dalam pemodelan regresi, variabel prediktor yang digunakan tidak boleh saling berkorelasi antara satu dengan yang lainnya. Terdapat dua cara untuk mengetahui adanya multikolinearitas pada variabel prediktor yang digunakan yaitu dengan melihat nilai koefisien korelasi antar variabel prediktor dan nilai *Variance Inflation Factor* (VIF) pada masing-masing prediktor.

Pada Tabel 4.2 menunjukkan bahwa tidak terdapat nilai koefisien korelasi antar variabel prediktor yang melebihi  $\pm 0,95$  sehingga dapat dikatakan bahwa pada variabel prediktor yang digunakan tidak memiliki adanya korelasi antar variabel yang bersifat multikolinearitas. Selanjutnya kasus multikolinearitas dapat dilihat dengan menggunakan nilai VIF. Pada Tabel 4.3 disajikan nilai VIF dari masing-masing prediktor.

**Tabel 4. 3** Nilai VIF Variabel Prediktor

Variabel	VIF
X <sub>1</sub>	2,150
X <sub>2</sub>	5,743
X <sub>3</sub>	4,201
X <sub>4</sub>	6,119
X <sub>5</sub>	5,621
X <sub>6</sub>	7,703
X <sub>7</sub>	2,795

Pada pengujian multikolinearitas dengan menggunakan nilai VIF, variabel dikatakan bermultikolinearitas jika nilai VIF lebih dari 10. Dikarenakan nilai VIF yang ditampilkan dalam Tabel 4.3 tidak ada yang melebihi nilai 10, maka dapat disimpulkan bahwa tidak ada kasus multikolinearitas antara variabel. Sehingga semua variabel prediktor yang digunakan dalam penelitian ini telah memenuhi asumsi non-multikolinearitas.

#### 4.3 Uji Overdispersi

Dataset yang digunakan pada penelitian ini terdiri dari 35 observasi dan merupakan data *count*. Seperti yang disebutkan oleh Cosul dan Famoye bahwa pada data *count* terkadang ditemukan kasus overdispersi. Maka dari itu pada analisis ini dilakukan pengujian overdispersi, dan jika hasilnya data mengalami overdispersi salah satu pemodelan yang dapat digunakan adalah regresi PIG.



Regresi PIG cocok untuk diaplikasikan pada data yang mengalami overdispersi dimana variansnya lebih besar dari *means*. Oleh karena itu data yang digunakan dalam penelitian ini diuji apakah mengalami overdispersi atau tidak dengan hipotesis sebagai berikut:

$$H_0 : \text{var}(Y) = \mu_i$$

$$H_1 : \text{var}(Y) = \mu_i + a.g(.)$$

Dengan menggunakan *package* AER pada *software* R, diperoleh nilai  $a = 15,22$  dan  $p\text{-value}$  sebesar 0,0001791 lebih kecil dari tingkat signifikansi 10% sehingga tolak  $H_0$  yang dapat disimpulkan bahwa varians tidak sama dengan rata-rata dan berarti bahwa data tersebut mengalami overdispersi. Cara kedua untuk menguji overdispersi dapat dilakukan dengan statistik Uji Khi Kuadrat Pearson. Ststistik uji ini dibagi dengan derajat bebas, apabila hasil pembagian menghasilkan nilai yang melebihi 1, maka dapat disimpulkan bahwa data mengalami overdispersi. Derajat bebas yang digunakan adalah 26, didapatkan dari pengurangan jumlah observasi yaitu 35 dengan jumlah koefisien regresi sebanyak 9. Hasil dari pengujian overdispersi menggunakan Uji Khi Kuadrat Pearson akan ditampilkan pada Tabel 4.4.

**Tabel 4. 4** Pengujian Overdispersi

$\bar{y}$	s	VT
41,91	26,45	567,561

Pada Tabel 4.4 dapat dilihat nilai *variance to mean ratio* (VT) sebesar 567,561. Dikarenakan hasil VT yang didapatkan melebihi nilai 1, maka dapat disimpulkan terjadi overdispersi, sehingga pemodelan dengan regresi PIG dapat dilakukan.

#### 4.4 Pemodelan Regresi Poisson Inverse Gaussian (PIG)

Regresi Poisson Inverse Gaussian (PIG) merupakan regresi yang diaplikasikan pada data yang mengalami overdispersi. Pemodelan regresi poisson inverse gaussian dapat dilakukan setelah pengujian asumsi multikolinearitas. Data yang digunakan untuk penerapan Regresi Poisson Inverse Gaussian adalah data jumlah kasus baru HIV di Provinsi Jawa Tengah tahun 2015. Sesuai dengan hasil pengujian yang sudah dilakukan, didapatkan bahwa data mengalami overdispersi dan tidak terdapat multikolinearitas, sehingga asumsi untuk melakukan pemodelan dengan regresi Poisson Inverse Gaussian (PIG) sudah terpenuhi.

Berdasarkan tujuh variabel yang signifikan pada model regresi PIG yaitu persentase penduduk miskin ( $X_1$ ), persentase penduduk dengan pendidikan tertinggi SLTA ( $X_2$ ), persentase pasangan usia subur yang sedang menggunakan kondom ( $X_3$ ), rasio jumlah tenaga kerja ( $X_4$ ), rasio fasilitas kesehatan ( $X_5$ ), persentase daerah perkotaan ( $X_6$ ), dan persentase penduduk usia 25-34 tahun ( $X_7$ ), menghasilkan banyak kombinasi model namun pada pembahasan hanya ditampilkan model yang sudah konvergen. Empat kombinasi kemungkinan model regresi PIG yang sudah konvergen, kemudian dicari model terbaiknya. Berikut merupakan empat kemungkinan model PIG yang sudah konvergen adalah sebagai berikut.

$$\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)$$

$$\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)$$

$$\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)$$

$$\mu = \exp(\beta_0 + \beta_3 x_3 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)$$

Sebelum membuat model, maka perlu dilakukan penaksiran parameter, dan pengujian parameter. Pengujian parameter

dilakukan secara serentak dan secara individu. Berikut merupakan estimasi parameter dari model-model yang mungkin menjadi model terbaik dalam regresi PIG yang ditunjukkan pada Tabel 4.5 sebagai berikut.

**Tabel 4. 5** Estimasi Parameter Kemungkinan Model Regresi PIG

Variabel dari Model	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	0,74202	-0,00650	-0,00031	0,39252
$X_1, X_3, X_4, X_5, X_6, X_7$	0,76540	-0,00870		0,39505
$X_1, X_3, X_5, X_6, X_7$	0,847088	-0,00752		0,38666
$X_3, X_5, X_6, X_7$	0,62996			0,38585

**Tabel 4. 5** (Lanjutan)

Variabel dari Model	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\tau$
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	-0,00031	-0,210	-0,021	0,452	-1,360
$X_1, X_3, X_4, X_5, X_6, X_7$	-0,00157	-0,209	-0,021	0,457	-1,324
$X_1, X_3, X_5, X_6, X_7$		-0,232	-0,022	0,452	-1,314
$X_3, X_5, X_6, X_7$		-0,230	-0,021	0,466	-1,316

Langkah selanjutnya setelah didapatkan nilai estimasi dari masing-masing perkiraan model yang ditampilkan pada Tabel 4.5 adalah pengujian hipotesis untuk regresi Poisson Inverse Gaussian.

#### 4.5 Pengujian Hipotesis

Pengujian hipotesis merupakan pengujian parameter yang dilakukan secara serentak dan secara individu untuk mengetahui signifikansi dari masing-masing parameter yang diperoleh.

##### 4.5.1 Pengujian Parameter Secara Serentak

Pengujian parameter secara serentak dilakukan pada kemungkinan model yang sesuai dengan model regresi Poisson Inverse Gaussian. Pengujian parameter secara serentak dapat dilihat dari nilai statistik G dengan hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

$$H_1 : \text{paling sedikit ada } \beta_i \neq 0 \text{ dengan } i=1,2,\dots,7$$

$$\alpha=0,1$$

Nilai statistik uji G yang diperoleh dari hasil regresi Poisson Inverse Gaussian akan dibandingkan dengan nilai  $\chi^2_{(\alpha, v)}$ . Pengujian parameter secara serentak dapat dilihat pada Tabel 4.6 sebagai berikut.

**Tabel 4. 6** Pengujian Parameter Regresi PIG Secara Serentak

Variabel dari Model	Statistik G	v	$\chi^2_{(\alpha, v)}$	Keputusan
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	299,146	27	36,741	Tolak $H_0$
$X_1, X_3, X_4, X_5, X_6, X_7$	299,1545	28	37,916	Tolak $H_0$
$X_1, X_3, X_5, X_6, X_7$	299,2379	29	39,087	Tolak $H_0$
$X_3, X_5, X_6, X_7$	299,3138	30	40,456	Tolak $H_0$

Tabel 4.6 menunjukkan bahwa dari semua kemungkinan model, didapatkan hasil statistik G lebih dari  $\chi^2_{(\alpha, v)}$  maka keputusannya adalah tolak  $H_0$ , yang berarti minimal terdapat satu parameter yang berpengaruh signifikan terhadap model. Untuk mengetahui variabel prediktor yang berpengaruh signifikan terhadap model, maka dilanjutkan pada pengujian parameter secara individu.

#### 4.5.2 Pengujian Parameter Secara Individu

Pengujian parameter secara individu digunakan untuk mencari variabel prediktor yang berpengaruh signifikan terhadap jumlah kasus baru HIV di Provinsi Jawa Tengah dengan hipotesis sebagai berikut.

- Parameter  $\beta$

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

$$\alpha = 0,1$$

- Parameter  $\tau$

$$H_0 : \tau = 0$$

$$H_1 : \tau \neq 0$$

$$\alpha = 0,1$$

Pengujian parameter secara individu dapat dilihat pada Tabel 4.7 sebagai berikut.

**Tabel 4. 7** Pengujian Parameter Regresi PIG Secara Individu

Variabel dari Model	Parameter Signifikan
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	$\beta_3, \beta_6, \beta_7, \tau$
$X_1, X_3, X_4, X_5, X_6, X_7$	$\beta_3, \beta_6, \beta_7, \tau$
$X_1, X_3, X_5, X_6, X_7$	$\beta_3, \beta_5, \beta_6, \beta_7, \tau$
$X_3, X_5, X_6, X_7$	$\beta_3, \beta_5, \beta_6, \beta_7, \tau$

Tabel 4.7 menunjukkan bahwa model dengan parameter yang signifikan adalah model yang memenuhi daerah kritis atau memiliki  $p$ -value kurang dari taraf signifikansi yaitu  $\alpha=0,1$  sehingga keputusan tolak  $H_0$  yang berarti parameter  $\beta_i$  berpengaruh signifikan terhadap model atau dengan melihat nilai  $|z_{hitung}|$  yang dibandingkan dengan  $z_{\alpha/2} = 1,64$ .  $P$ -value dan  $|z_{hitung}|$  dari semua model dapat dilihat pada Lampiran 5.

#### 4.6 Pemilihan Model Terbaik

Pada penelitian ini pemodelan menggunakan *package* *gamlss* yang tersedia pada *software* R. Pemilihan model regresi PIG terbaik dalam penelitian ini dilakukan dengan metode *allpossible elimination* yang bertujuan untuk mendapatkan model PIG dengan variabel-variabel prediktor yang signifikan. Metode *allpossible elimination* dilakukan berdasarkan nilai *Akaike Information Criterion* (AIC). Metode ini mengeliminasi beberapa variabel prediktor sehingga dapat menghasilkan nilai AIC yang lebih kecil. Model regresi Poisson Inverse Gaussian terbaik yang dipilih adalah dengan model yang memiliki nilai AIC terkecil. Dengan mengeluarkan variabel  $X_1$ ,  $X_2$ , dan  $X_4$  dari model akan menghasilkan model regresi dengan nilai AIC yang lebih kecil. Variabel-variabel prediktor yang signifikan untuk pemodelan

regresi Poisson Inverse Gaussian adalah persentase PUS yang menggunakan kondom ( $X_3$ ), rasio jumlah fasilitas kesehatan ( $X_5$ ), persentase daerah perkotaan ( $X_6$ ), dan persentase penduduk usia 25-34 tahun ( $X_7$ ). Pada Tabel 4.8 disajikan nilai AIC model Regresi Poisson Inverse Gaussian.

**Tabel 4. 8** Nilai AIC dari Model Regresi Poisson Inverse Gaussian

Model	AIC
$X_1, X_2, X_3, X_4, X_5, X_6, X_7$	317,146
$X_1, X_3, X_4, X_5, X_6, X_7$	315,1545
$X_1, X_3, X_5, X_6, X_7$	313,2379
$X_3, X_5, X_6, X_7$	311,3138

Pada Tabel 4.8 terlihat bahwa nilai AIC yang paling kecil yaitu model yang mengandung variabel  $X_3$ ,  $X_5$ ,  $X_6$ , dan  $X_7$ . Oleh karena itu model ke empat yang dipilih sebagai model terbaik. Berdasarkan model tersebut, dengan menggunakan *package* gamlss yang tersedia pada *software* R didapatkan hasil estimasi parameter model Regresi Poisson Inverse Gaussian yang disajikan pada Tabel 4.9.

**Tabel 4. 9** Penaksiran Parameter Model Regresi PIG pada Jumlah Kasus Baru HIV di Provinsi Jawa Tengah Tahun 2015

Parameter	Taksiran	Standard Error	Z Hitung	<i>P-value</i>
$\beta_0$	0,629965	1,788147	0,352	0,72718
$\beta_3$	0,385849	0,108118	3,569	0,00127*
$\beta_5$	-0,23026	0,116140	-1,983	0,05695*
$\beta_6$	-0,02185	0,007278	-3,003	0,00546*
$\beta_7$	0,466977	0,212157	2,201	0,03584*
$\tau$	-1,3161	0,3079	-4,274	0,000189*

\*) Signifikan dengan taraf signifikansi  $\alpha = 0,1$

Berdasarkan Tabel 4.9 diperoleh model regresi PIG sebagai berikut.

$$\hat{\mu} = \exp(a + bX_3 - cX_5 - dX_6 + eX_7)$$

dimana

$$a = 0,629965$$

$$b = 0,385849$$

$$c = 0,230265$$

$$d = 0,021858$$

$$e = 0,466977$$

Berdasarkan model di atas, maka dapat diinterpretasikan sebagai berikut ini. Setiap penambahan 1 persen variabel  $X_3$  maka akan melipatgandakan rata-rata variabel respon  $Y$  sebesar  $\exp(0,385849)=1,47086$  kali dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 persen dari persentase PUS yang menggunakan kondom maka akan sebanding dengan kenaikan rata-rata jumlah kasus baru HIV sebesar 1,47086 kali dari rata-ratanya semula bila variabel lain tetap. Hal ini berbanding terbalik jika dibandingkan dengan teori. Alasan mengapa hal ini dapat terjadi adalah dikarenakan tujuan dari PUS untuk menggunakan kondom adalah untuk perihal reproduksi bukan untuk pencegahan HIV, selain itu penggunaan kondom tidak menutup kemungkinan digunakan oleh seseorang yang sudah terkena HIV sehingga kondom digunakan untuk pencegahan penyebaran virus tersebut kepada pasangan. Oleh karena beberapa alasan tersebut maka interpretasi pada model ini tidak dapat dilihat sebagai sebab akibat.

Untuk variabel  $X_5$  dinyatakan bahwa setiap penambahan 1 persen variabel  $X_5$  maka akan melipatgandakan rata-rata variabel respon  $Y$  sebesar  $\exp(-0,230265) = 0,79432$  kali dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 rasio persentase fasilitas kesehatan maka akan sebanding dengan penurunan rata-rata jumlah kasus baru HIV sebesar 0,79432 kali dari rata-ratanya semula bila variabel lain tetap.

Untuk variabel  $X_6$  dinyatakan bahwa setiap penambahan 1 persen variabel  $X_6$  maka akan melipatgandakan rata-rata variabel respon  $Y$  sebesar  $\exp(-0,021858)=0,97838$  kali dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain,

penambahan 1 persen dari persentase daerah perkotaan akan sebanding dengan penurunan rata-rata jumlah kasus baru HIV sebesar 0,97838 kali dari rata-ratanya semula bila variabel lain tetap.

Untuk variabel  $X_7$  dinyatakan bahwa Setiap penambahan 1 persen variabel  $X_7$  maka akan melipatgandakan rata-rata variabel respon  $Y$  sebesar  $\exp(0,466977)=1,595164714$  kali dari rata-rata variabel respon semula bila variabel lain tetap. Dengan kata lain, penambahan 1 persen dari persentase penduduk usia 25-34 tahun maka akan sebanding dengan kenaikan rata-rata jumlah kasus baru HIV sebesar 1,595164714 kali dari rata-ratanya semula bila variabel lain tetap.

Masalah keterbatasan data yang tersedia menyebabkan beberapa interpretasi dari model yang terbentuk tidak sesuai dengan teori yang berlaku tentang HIV, sehingga interpretasi tersebut tidak dapat dilihat sebagai hubungan sebab akibat. Data jumlah kasus baru HIV yang digunakan dalam penelitian ini merupakan data yang berasal dari hasil laporan klinik-klinik VCT di Provinsi Jawa Tengah dimana pasien datang secara sukarela untuk memeriksakan diri ke klinik tersebut.



*(Halaman ini sengaja dikosongkan)*

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan dapat diambil kesimpulan sebagai berikut:

1. Pada tahun 2015 jumlah kasus baru HIV di Provinsi Jawa Tengah mengalami peningkatan sebesar 115 kasus dibandingkan tahun sebelumnya. Jumlah kasus baru HIV tertinggi adalah di Kota Semarang. Setelah dilakukan uji overdispersi pada data jumlah kasus baru HIV di Provinsi Jawa Tengah tahun 2015 dinyatakan bahwa data mengalami overdispersi. Oleh karena itu penelitian dengan menggunakan metode Regresi Poisson Inverse Gaussian dapat dilakukan. Model Regresi Poisson Inverse Gaussian (PIG) yang terbentuk adalah sebagai berikut:

$$\hat{\mu} = \exp(a + bX_3 - cX_5 - dX_6 + eX_7)$$

dimana

$$a = 0,629965$$

$$b = 0,385849$$

$$c = 0,230265$$

$$d = 0,021858$$

$$e = 0,466977$$

2. Berdasarkan model Regresi Poisson Inverse Gaussian yang terbentuk maka didapatkan bahwa faktor-faktor yang berpengaruh signifikan terhadap jumlah kasus baru HIV di Provinsi Jawa Tengah pada tahun 2015 yaitu persentase PUS yang menggunakan kondom ( $X_3$ ), rasio fasilitas kesehatan ( $X_5$ ), persentase daerah perkotaan ( $X_6$ ), dan persentase penduduk usia 25-34 tahun ( $X_7$ ). Berikut hasil pemodelan dari regresi Poisson Inverse Gaussian.

## 5.2 Saran

Saran yang diberikan berdasarkan hasil analisis yang telah dilakukan sebelumnya adalah sebagai berikut.

1. Berdasarkan hasil penelitian ini, saran yang diberikan kepada Pemerintah Daerah (Pemda) dan dinas-dinas dalam bidang kesehatan Provinsi Jawa Tengah adalah meningkatkan sosialisasi tentang virus HIV serta penyebarannya kepada berbagai lapisan masyarakat terutama pada masyarakat yang rentan tertular virus tersebut.
2. Pada penelitian selanjutnya diharapkan dapat dilakukan pengembangan metode pemodelan regresi Poisson Inverse Gaussian dengan memasukkan efek spasial sehingga mendapatkan hasil perkiraan yang lebih baik.

## DAFTAR PUSTAKA

- Akaike, H. (1978). *A Bayesian Analysis of The Minimum AIC Procedure*. Annals of the Institute of Statistical Mathematics, Part A Hal. 914. <http://www.ism.ac.jp/editsec/aism/pdf/> Tanggal Akses: 1 Oktober 2016.
- BPS Jawa Tengah. (2016). *Statistik Sosial dan Kependudukan Jawa Tengah Hasil Susenas 2015*. Semarang: BPS Jawa Tengah.
- BPS Republik Indonesia. (2013). *Survei Demografi daan Kesehatan Indonesia Tahun 2012*. Jakarta: BPS Republik Indonesia.
- Cameron, A. C. dan Trivedi, P. K. (1990), "Regression-Based Test for Overdispersion In The Poisson Model", *Journal of Econometrics*, Vol. 46, No.1, hal. 347-346.
- CDC. (2016). *HIV/AIDS*. <http://www.cdc.gov/hiv/statistics/index.html>. Tanggal Akses: 5 Oktober 2016.
- Cole. T. J. Dan Green, P. J. (1992), "Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood", *Statistics in Medicine*, Vol. 11, hal 1305-1319.
- Consul, P.C. dan Famoye, F. (1992). "Generalized Poisson Regression Model", *Commonication in Statistics – Theory and Methods*. Vol. 21, No.1, hal. 89-109
- Damar I. (2015). *Jumlah Kasus HIV & AIDS di Indonesia Meningkat*. <http://news.metrotvnews.com/read/2015/11/30/196222/jumlah-kasus-hiv-aids-di-indonesia-meningkat>. Tanggal Akses: 5 Oktober 2016.
- Dinas Kesehatan Provinsi Jawa Tengah. (2013). *Profil Kesehatan Provinsi Jawa Tengah Tahun 2013*. Semarang: Dinkes Jateng.
- Dinas Kesehatan Provinsi Jawa Tengah. (2016). *Buku Saku Kesehatan Provinsi Jawa Tengah Tahun 2015*. Semarang: Dinkes Jateng.
- De Jong, P. dan Heller, G.Z. (2008), "*Generalized Linear Model for Insurance Data*", 1<sup>st</sup> edition, Cambridge University, Press, New York.

- Draper, N. dan Smith, H. (1992). *Analisis Regresi Terapan*. Jakarta: Gramedia.
- Karlis, D. dan Nikolouloupoulos, E. (2005), “Mixed Poisson Distribution”, *International Statical Review*, Vol. 73, No.1 hal. 35-58.
- Karlis, D. dan Xekalaki, E. (2000), “ A Simulation Comparison of Several Produces for Testing the Poisson Assumption”, *The Statistican*, Vol. 49, No.3, hal. 355-382.
- Myers, R.H., Montgomery, D.C., Vining, G.G., dan Robinson, T.J. (2010). *Generalized Linier Models with Aplication in Engineering and Sciences*. John Wiley and Sons, Inc., Publication. Canada.
- Purnamasari, I. (2016). *Penaksiran Parameter Dan Statistik Uji Dalam Model Regresi Geographically Weighted Poisson Inverse Gaussian*. Tesis. Mahasiswa Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya.
- Ratnasari, N. T. dan Purhadi. (2013). *Pemodelan Faktor Yang Mempengaruhi Jumlah HIV Dan AIDS Provinsi Jawa Timur Menggunakan Regresi Poisson Bivariat*. Institut Teknologi Sepuluh Nopember, Surabaya.
- Rigby, R. A. dan Stasinopoulus, D. M. (1996), “A Semi-parametric Additive Model for variance Heterogeneity”, *Statistics and Computing*, Vol. 6, hal. 57-65.
- Setiawan, dan Kusrini, D. E. (2010). *Ekonometrika*. Yogyakarta: C.V. Andi Offset.
- Shoukri, M.M., Asyali, M. H., Vandorp, R. dan Kelton, R. (2004) “ The Poisson Inverse Gaussian Regression Model in the Analysis of Cluster Counts Data”, *Journal of Data Science*, Vol.2, No. 2, hal 17-32.
- Stasinopoulus, D. M. dan Rigby, R. A. (2007a), “Generalized Additive Models for Location Scale and Shape”, *Journal of Statistical Software*, Vol.23, hal. 1-46.
- Stasinopoulus, D. M. dan Rigby, R. A. (2007b), “Generalized Additive Models for Location Scale and Shape

- (GAMLSS) in R”, *Journal of Statistical Software*, Vol.20, hal. 1-64.
- Triola, M. (2010). “*Elementary Statistics*”, 11<sup>th</sup> edition, Addison Wesley, Press., New York.
- UNAIDS. (2016). *Global AIDS Update 2016*. Geneva: WHO.
- Walpole, R.E. (1995). *Pengantar Metode Statistika*. Ahli Bahasa: Ir. Bambang Sumantri, Jakarta : PT Gramedia Pustaka Utama.
- WHO. (2016). *Data and statistics HIV/AIDS*. <http://www.who.int/hiv/data/en/>. Tanggal Akses : 5 Oktober 2016.
- Whiteside, A. (2002), “Poverty and HIV/AIDS in Africa”, *Third World Quarterly*, Vol.23, No.2, hal. 313-332.
- Wilmot, G. E. (1987), “The Poisson-Inverse Gaussian Distribution as An Alternative to the Negative Binomial”, *Scandinavian Actuarial Jurnal*, Vol. 3, No. 4, hal 113-127.
- Widiari, S. M. (2016). *Penaksiran Parameter Dan Statistik Uji Dalam Model Regresi Poisson Inverse Gaussian (PIG) Studi Kasus: Jumlah Kasus Baru HIV di Provinsi Jawa Timur Tahun 2013*. Tesis. Mahasiswa Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya.
- Zha, L., Lord, D. dan Zou, Y. (2014), “The Poisson Inverse Gaussian (PIG) Generalized Linear Regression Model for Analyzing Motor Vehicle Crash Data”, *Journal of Transportation Safety and Security*, DOI: 10.1080/19439962.2014.977502.
- Zhu, R. dan Joe, H. (2009), “Modelling heavy-tailed count data using a generalized Poisson-inverse Gaussian family”, *Statistics and Probability Letters*, DOI: 10.1016/j.spl.2009.04.011

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN

### Lampiran 1 Data Jumlah Kasus Baru HIV serta Faktor-Faktor yang Diduga Mempengaruhi di Provinsi Jawa Tengah Tahun 2015

Kab/Kota	Y	X1	X2	X3	X4	X5	X6	X7
Kab. Cilacap	60	14.39	7.62	2.35	5.13	2.83	5.28	7.055
Kab. Banyumas	61	17.52	8.62	2.41	17.48	3.79	9.06	7.03
Kab. Purbalingga	9	19.7	7.76	1.40	8.68	3.01	6.28	7.135
Kab. Banjarnegara	41	18.37	7.19	2.06	10.20	4.21	4.32	7.16
Kab. Kebumen	61	20.44	8.47	2.22	18.14	3.88	2.39	6.78
Kab. Purworejo	16	14.27	11.76	2.42	18.72	4.79	5.06	6.415
Kab. Wonosobo	45	21.45	6.68	0.95	14.03	3.60	10.94	7.285
Kab. Magelang	14	13.07	11.65	1.67	13.57	2.65	1.34	6.665
Kab. Boyolali	59	12.45	12.93	0.78	21.17	4.05	2.25	6.94
Kab. Klaten	44	14.89	14.4	2.05	21.32	3.97	2.49	7.13
Kab. Sukoharjo	40	9.26	17.4	1.22	52.53	2.20	10.18	7.95
Kab. Wonogiri	7	12.98	8.28	1.79	16.54	4.43	14.63	5.82
Kab. Karanganyar	20	12.46	14.41	2.00	31.65	3.39	8.47	7.515
Kab. Sragen	62	14.86	10.3	1.81	28.10	3.98	5.77	7.025
Kab. Grobogan	65	13.68	7.58	0.95	15.91	2.74	2.50	7.5
Kab. Blora	29	13.52	8.54	1.50	13.14	3.76	8.14	7.16
Kab. Rembang	28	19.28	10.81	0.69	13.08	2.91	2.38	8.08
Kab. Pati	47	11.95	12.48	1.27	14.84	3.08	1.23	7.53
Kab. Kudus	31	7.73	17.13	1.76	27.31	3.49	6.82	8.21
Kab. Jepara	82	8.5	13.01	1.98	15.90	2.44	5.64	7.835
Kab. Demak	61	14.44	13.63	1.72	10.02	2.68	2.41	8.025
Kab. Semarang	45	8.15	10.22	0.82	27.08	3.10	11.49	7.985



**Lampiran 1** Lanjutan

<b>Kab/Kota</b>	<b>Y</b>	<b>X1</b>	<b>X2</b>	<b>X3</b>	<b>X4</b>	<b>X5</b>	<b>X6</b>	<b>X7</b>
Kab. Temanggung	32	11.76	6.74	2.26	29.10	3.75	7.96	7.005
Kab. Kendal	38	11.62	11.84	0.35	17.62	3.61	6.99	7.095
Kab. Batang	99	11.27	9.51	3.11	14.53	3.09	3.63	7.685
Kab. Pekalongan	31	12.84	8.73	1.25	24.60	3.32	4.56	7.71
Kab. Pemalang	72	18.3	7.92	1.89	10.63	2.25	4.95	7.24
Kab. Tegal	52	10.09	9.87	0.68	11.86	2.53	2.09	7.03
Kab. Brebes	37	19.79	6.87	1.51	9.43	2.75	1.68	7.49
Kota Magelang	7	9.05	23.32	5.97	222.72	9.94	100.00	7.195
Kota Surakarta	29	10.89	25.23	7.41	130.03	5.86	100.00	8.135
Kota Salatiga	18	5.8	22.59	5.88	100.64	7.07	100.00	8.665
Kota Semarang	116	4.97	24.98	5.50	60.19	3.76	100.00	9.09
Kota Pekalongan	6	8.09	18.23	4.08	68.80	7.42	100.00	8.335
Kota Tegal	3	8.26	15.2	2.14	71.52	4.88	100.00	8.47

Keterangan:

Y = Jumlah kasus baru HIV di Provinsi Jawa Tengah Tahun 2015.

X<sub>1</sub> = Persentase penduduk miskin di kabupaten/kota Provinsi Jawa Tengah tahun 2015

X<sub>2</sub> = Persentase penduduk pendidikan tertinggi SLTA di kabupaten/kota Provinsi Jawa Tengah tahun 2015

X<sub>3</sub> = Persentase PUS yang sedang menggunakan alat KB kondom di kabupaten/kota Provinsi Jawa Tengah tahun 2015

X<sub>4</sub> = Rasio jumlah tenaga kesehatan per 100.000 penduduk di kabupaten/kota Provinsi Jawa Tengah tahun 2015

X<sub>5</sub> = Rasio jumlah fasilitas kesehatan per 100.000 penduduk di kabupaten/kota Provinsi Jawa Tengah tahun 2015

X<sub>6</sub> = Persentase daerah perkotaan di kabupaten/kota Provinsi Jawa Tengah tahun 2015

X<sub>7</sub> = Persentase penduduk usia 25-34 tahun di kabupaten/kota Provinsi Jawa Tengah tahun 2015

**Lampiran 2** Program R pada Uji Overdispersi

```
library(AER)

data<-read.csv("F:/bis.csv",header=TRUE)

Y=data[,1]

X1=data[,2]

X2=data[,3]

X3=data[,4]

X4=data[,5]

X5=data[,6]

X6=data[,7]

X7=data[,8]

rd<-glm(Y~1,data=data,family=poisson)

dispersiontest(rd,trafo=1)
```

**Lampiran 3** Hasil Program R pada Uji Overdispersi

## Overdispersion test

data: rd

$z = 3.5691$ , p-value = 0.0001791

alternative hypothesis: true alpha is greater than 0

sample estimates:

alpha

15.22

#### **Lampiran 4** Program R pada Regresi Poisson Inverse Gaussian (PIG)

```
library(gamlss)
data<-read.csv("F:/bis.csv",header=TRUE)
Y=data[,1]
X1=data[,2]
X2=data[,3]
X3=data[,4]
X4=data[,5]
X5=data[,6]
X6=data[,7]
X7=data[,8]

step(gamlss(Y~X1+X2+X3+X4+X5+X6+X7,family=PIG,data=
  data,method=mixed(10,50),direction="backward"))

summary(m1<gamlss(Y~X3+X5+X6+X7,family=PIG,data=data
  ,method=mixed(10,50),direction="backward"))

summary(m2<gamlss(Y~X1+X2+X3+X4+X5+X6+X7,family=
  PIG,data=data,method=mixed(10,50),direction="backward"))

summary(m3<gamlss(Y~X1+X3+X4+X5+X6+X7,family=PIG,
  data=data,method=mixed(10,50),direction="backward"))

summary(m4<gamlss(Y~X1+X3+X5+X6+X7,family=PIG,data
  =data,method=mixed(10,50),direction="backward"))
```

### Lampiran 5 Hasil Program R pada Regresi Poisson Inverse Gaussian (PIG)

```
>step(gamlss(Y~X1+X2+X3+X4+X5+X6+X7,family=PIG,data=
  data,method=mixed(10,50),direction="allpossible"))
```

GAMLSS-RS iteration 1: Global Deviance = 299.2975

GAMLSS-RS iteration 2: Global Deviance = 299.1221

GAMLSS-RS iteration 3: Global Deviance = 299.1189

GAMLSS-RS iteration 4: Global Deviance = 299.1194

GAMLSS-CG iteration 1: Global Deviance = 299.1277

GAMLSS-CG iteration 2: Global Deviance = 299.2456

GAMLSS-CG iteration 3: Global Deviance = 299.1467

GAMLSS-CG iteration 4: Global Deviance = 299.146

Start: AIC=317.15

$Y \sim X1 + X2 + X3 + X4 + X5 + X6 + X7$

trying - X1

trying - X2

trying - X3

Model with term X3 has failed

trying - X4

trying - X5

trying - X6

trying - X7

	Df	AIC	LRT	Pr(Chi)
- X2	1	315.15	0.0085	0.926445
- X4	1	315.16	0.0136	0.907143
- X1	1	315.26	0.1171	0.732156
<none>		317.15		
- X5	1	317.38	2.2322	0.135164
- X7	1	319.35	4.2015	0.040388 *
- X6	1	323.95	8.8068	0.003001 **
- X3				

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

GAMLSS-RS iteration 1: Global Deviance = 299.3297

GAMLSS-RS iteration 2: Global Deviance = 299.1572  
 GAMLSS-RS iteration 3: Global Deviance = 299.155  
 GAMLSS-RS iteration 4: Global Deviance = 299.1545  
 GAMLSS-CG iteration 1: Global Deviance = 299.1545

Step: AIC=315.15

$Y \sim X1 + X3 + X4 + X5 + X6 + X7$

trying - X1

trying - X3

trying - X4

trying - X5

Model with term X5 has failed

trying - X6

trying - X7

	Df	AIC	LRT	Pr(Chi)
- X4	1	313.24	0.0834	0.7727363
- X1	1	313.26	0.1070	0.7435772
<none>		315.15		
- X7	1	317.56	4.4047	0.0358409 *
- X6	1	321.99	8.8324	0.0029593 **
- X3	1	324.77	11.6195	0.0006526 ***
- X5				

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

GAMLSS-RS iteration 1: Global Deviance = 299.4038  
 GAMLSS-RS iteration 2: Global Deviance = 299.2409  
 GAMLSS-RS iteration 3: Global Deviance = 299.2392  
 GAMLSS-RS iteration 4: Global Deviance = 299.2379  
 GAMLSS-RS iteration 5: Global Deviance = 299.238  
 GAMLSS-CG iteration 1: Global Deviance = 299.2404  
 GAMLSS-CG iteration 2: Global Deviance = 299.275  
 GAMLSS-CG iteration 3: Global Deviance = 299.2399  
 GAMLSS-CG iteration 4: Global Deviance = 299.2379  
 GAMLSS-CG iteration 5: Global Deviance = 299.2379

Step: AIC=313.24

Y ~ X1 + X3 + X5 + X6 + X7

trying - X1

trying - X3

trying - X5

trying - X6

trying - X7

	Df	AIC	LRT	Pr(Chi)
- X1	1	311.31	0.0759	0.7829875
<none>		313.24		
- X5	1	314.91	3.6716	0.0553458 .
- X7	1	315.56	4.3253	0.0375504 *
- X6	1	320.37	9.1292	0.0025156 **
- X3	1	323.17	11.9301	0.0005523 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

GAMLSS-RS iteration 1: Global Deviance = 299.4784

GAMLSS-RS iteration 2: Global Deviance = 299.3157

GAMLSS-RS iteration 3: Global Deviance = 299.3147

GAMLSS-RS iteration 4: Global Deviance = 299.3136

GAMLSS-RS iteration 5: Global Deviance = 299.3136

GAMLSS-CG iteration 1: Global Deviance = 299.3138

Step: AIC=311.31

Y ~ X3 + X5 + X6 + X7

trying - X3

trying - X5

trying - X6

trying - X7

	Df	AIC	LRT	Pr(Chi)
<none>		311.31		
- X5	1	312.98	3.6710	0.055367 .

```
- X7  1 314.16  4.8440 0.027742 *
- X6  1 318.50  9.1902 0.002433 **
- X3  1 321.19 11.8715 0.000570 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Family: c("PIG", "Poisson.Inverse.Gaussian")
```

```
Fitting method: mixed(10, 50)
```

```
Call:  gamlss(formula = Y ~ X3 + X5 + X6 + X7, family = PIG,
             data = data,
             method = mixed(10, 50), direction = "allpossible")
```

```
Mu Coefficients:
```

```
(Intercept)      X3      X5      X6      X7
    0.62996    0.38585   -0.23026  -0.02186    0.46698
```

```
Sigma Coefficients:
```

```
(Intercept)
    -1.316
```

```
Degrees of Freedom for the fit: 6 Residual Deg. of Freedom  29
```

```
Global Deviance:  299.314
```

```
AIC:  311.314
```

```
SBC:  320.646
```

```
> summary(m1<-
```

```
  gamlss(Y~X1+X2+X3+X4+X5+X6+X7,family=PIG,data=d
    ata,method=mixed(10,50),direction="allpossible"))
```

```
GAMLSS-RS iteration 1: Global Deviance = 299.2975
```

```
GAMLSS-RS iteration 2: Global Deviance = 299.1221
```

```
GAMLSS-RS iteration 3: Global Deviance = 299.1189
```

```
GAMLSS-RS iteration 4: Global Deviance = 299.1194
```

```
GAMLSS-CG iteration 1: Global Deviance = 299.1277
```

```
GAMLSS-CG iteration 2: Global Deviance = 299.2456
```

```
GAMLSS-CG iteration 3: Global Deviance = 299.1467
```

```
GAMLSS-CG iteration 4: Global Deviance = 299.146
```



Family: c("PIG", "Poisson.Inverse.Gaussian")

Call:

```
gamlss(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7,
       family = PIG,
       data = data, method = mixed(10, 50), direction = "allpossible")
```

Fitting method: mixed(10, 50)

-----  
Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value
(Intercept)	0.7420213	1.9208321	0.386
X1	-0.0065026	0.0299336	-0.217
X2	-0.0003120	0.0397372	-0.008
X3	0.3925206	0.1152458	3.406
X4	-0.0003149	0.0056045	-0.056
X5	-0.2104764	0.1356350	-1.552
X6	-0.0218928	0.0073798	-2.967
X7	0.4522741	0.2262483	1.999

Pr(>|t|)

(Intercept)	0.70242
X1	0.82972
X2	0.99379
X3	0.00215 **
X4	0.95563
X5	0.13280
X6	0.00638 **
X7	0.05616 .

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----

Sigma link function: log

Sigma Coefficients:

Estimate Std. Error t value

(Intercept) -1.3600 0.3061 -4.443

Pr(>|t)

(Intercept) 0.000146 \*\*\*

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----  
No. of observations in the fit: 35

Degrees of Freedom for the fit: 9

Residual Deg. of Freedom: 26

at cycle: 4

Global Deviance: 299.146

AIC: 317.146

SBC: 331.1441

\*\*\*\*\*

\*\*\*\*\*

> summary(m2<-

gamlss(Y~X1+X3+X4+X5+X6+X7,family=PIG,data=data,  
method=mixed(10,50),direction="allpossible"))

GAMLSS-RS iteration 1: Global Deviance = 299.3297

GAMLSS-RS iteration 2: Global Deviance = 299.1572

GAMLSS-RS iteration 3: Global Deviance = 299.155

GAMLSS-RS iteration 4: Global Deviance = 299.1545

GAMLSS-CG iteration 1: Global Deviance = 299.1545

\*\*\*\*\*

\*\*\*\*\*

Family: c("PIG", "Poisson.Inverse.Gaussian")

Call:

```
gamlss(formula = Y ~ X1 + X3 + X4 + X5 + X6 + X7, family =
      PIG,
      data = data, method = mixed(10, 50), direction = "allpossible")
```

Fitting method: mixed(10, 50)

-----

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value
(Intercept)	0.765401	1.943419	0.394
X1	-0.008705	0.026557	-0.328
X3	0.395051	0.111357	3.548
X4	-0.001575	0.005319	-0.296
X5	-0.209334	0.137353	-1.524
X6	-0.021964	0.007486	-2.934
X7	0.457743	0.217963	2.100

Pr(>|t|)

(Intercept) 0.69679

X1 0.74559

X3 0.00145 \*\*

X4 0.76941

X5 0.13912

X6 0.00675 \*\*

X7 0.04520 \*

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-1.3242	0.3097	-4.276

Pr(>|t|)

(Intercept) 0.000213 \*\*\*

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----  
No. of observations in the fit: 35

Degrees of Freedom for the fit: 8

Residual Deg. of Freedom: 27  
at cycle: 1

Global Deviance: 299.1545

AIC: 315.1545

SBC: 327.5973

\*\*\*\*\*

\*\*\*\*\*

> summary(m3<-

gamlss(Y~X1+X3+X5+X6+X7,family=PIG,data=data,method=mixed(10,50),direction="allpossible"))

GAMLSS-RS iteration 1: Global Deviance = 299.4038

GAMLSS-RS iteration 2: Global Deviance = 299.2409

GAMLSS-RS iteration 3: Global Deviance = 299.2392

GAMLSS-RS iteration 4: Global Deviance = 299.2379

GAMLSS-RS iteration 5: Global Deviance = 299.238

GAMLSS-CG iteration 1: Global Deviance = 299.2404

GAMLSS-CG iteration 2: Global Deviance = 299.275

GAMLSS-CG iteration 3: Global Deviance = 299.2399

GAMLSS-CG iteration 4: Global Deviance = 299.2379

GAMLSS-CG iteration 5: Global Deviance = 299.2379

\*\*\*\*\*

\*\*\*\*\*

Family: c("PIG", "Poisson.Inverse.Gaussian")

Call:

gamlss(formula = Y ~ X1 + X3 + X5 + X6 + X7, family = PIG,  
data = data,

method = mixed(10, 50), direction = "allpossible")

Fitting method: mixed(10, 50)  
-----

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value
(Intercept)	0.847088	1.936229	0.437
X1	-0.007256	0.026163	-0.277
X3	0.386664	0.108526	3.563
X5	-0.232773	0.116197	-2.003
X6	-0.022231	0.007440	-2.988
X7	0.452520	0.218349	2.072

	Pr(> t )
(Intercept)	0.66511
X1	0.78357
X3	0.00134 **
X5	0.05491 .
X6	0.00579 **
X7	0.04754 *

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-1.3140	0.3078	-4.269

Pr(>|t|)

(Intercept) 0.000203 \*\*\*

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----

No. of observations in the fit: 35

Degrees of Freedom for the fit: 7

Residual Deg. of Freedom: 28

at cycle: 5

Global Deviance: 299.2379

AIC: 313.2379

SBC: 324.1254

\*\*\*\*\*

\*\*\*\*\*

```
> summary(m4<-
  gamlss(Y~X3+X5+X6+X7,family=PIG,data=data,method=
    mixed(10,50),direction="allposible"))
```

GAMLSS-RS iteration 1: Global Deviance = 299.4784

GAMLSS-RS iteration 2: Global Deviance = 299.3157

GAMLSS-RS iteration 3: Global Deviance = 299.3147

GAMLSS-RS iteration 4: Global Deviance = 299.3136

GAMLSS-RS iteration 5: Global Deviance = 299.3136

GAMLSS-CG iteration 1: Global Deviance = 299.3138

\*\*\*\*\*

\*\*\*\*\*

Family: c("PIG", "Poisson.Inverse.Gaussian")

Call:

```
gamlss(formula = Y ~ X3 + X5 + X6 + X7, family = PIG, data =
  data,
```

```
  method = mixed(10, 50), direction = "allposible")
```

```
Fitting method: mixed(10, 50)
```

-----  
Mu link function: log

Mu Coefficients:

Estimate Std. Error t value

(Intercept) 0.629965 1.788147 0.352

X3 0.385849 0.108118 3.569

X5 -0.230265 0.116140 -1.983

X6 -0.021858 0.007278 -3.003

X7 0.466977 0.212157 2.201

Pr(>|t|)

(Intercept) 0.72716

X3 0.00127 \*\*

X5        0.05695 .  
 X6        0.00546 \*\*  
 X7        0.03584 \*

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----  
 Sigma link function: log

Sigma Coefficients:

Estimate Std. Error t value

(Intercept) -1.3161    0.3079 -4.274

Pr(>|t|)

(Intercept) 0.000189 \*\*\*

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----  
 No. of observations in the fit: 35

Degrees of Freedom for the fit: 6

Residual Deg. of Freedom: 29

at cycle: 1

Global Deviance:    299.3138

AIC:    311.3138

SBC:    320.6459

\*\*\*\*\*

\*\*\*\*\*

**Lampiran 6 Surat Pernyataan Penambilan Data****SURAT PERNYATAAN**

Saya yang bertanda tangan di bawah ini, mahasiswa jurusan  
Statistika FMIPA ITS:

Nama : Andriana Yoshinta H.

NRP : 1313100015

menyatakan bahwa data yang saya gunakan dalam Tugas Akhir/Thesis ini merupakan data sekunder yang diambil dari penelitian/buku/Tugas Akhir/Thesis/publikasi lainnya yaitu:

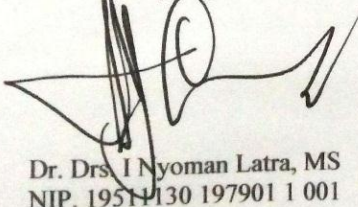
Sumber : Buku Saku Kesehatan Tahun 2015 Provinsi Jawa Tengah, BPS Provinsi Jawa Tengah

Keterangan : Data jumlah kasus baru HIV beserta faktor-faktor yang mempengaruhi diambil dari buku saku kesehatan 2015 dan hasil susenas yang di publish BPS Jawa Tengah

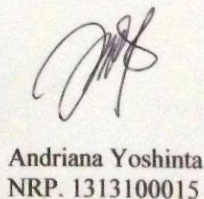
Surat pernyataan ini dibuat dengan sebenar-benarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Mengetahui,  
Pembimbing Tugas Akhir

Surabaya, 23 Januari 2017



Dr. Drs. I Nyoman Latra, MS  
NIP. 19511130 197901 1 001



Andriana Yoshinta  
NRP. 1313100015



*(Halaman ini sengaja dikosongkan)*

## BIODATA PENULIS



Penulis bernama Andriana Yoshinta Herindrawati yang biasa dipanggil “Yoshi” lahir di Jombang, 12 Agustus 1994. Penulis bertempat tinggal di Desa Sengon Kecamatan Jombang, Kabupaten Jombang. Pendidikan formal yang telah diselesaikan adalah di TK Trisula II, SDN Jombatan IV (2001-2007), SMP Negeri 2 Jombang (2007-2010), dan SMA Negeri 2 Jombang (2010-2013). Setelah lulus dari SMA penulis diterima di Jurusan

Statistika ITS dengan NRP 1313100015.

Segala kritik dan saran yang membangun akan diterima oleh penulis untuk perbaikan kedepannya. Jika ada keperluan atau ingin berdiskusi dengan penulis dapat dihubungi melalui email [andrianayoshinta7@gmail.com](mailto:andrianayoshinta7@gmail.com). Dengan segala ketulusan hati penulis mengucapkan terimakasih kepada pembaca.